REMARKS

In the Final Office Action dated May 30, 2008, claims 1-8, 13 and 16-83 are pending. Claims 1, 5, 13, 16, 32-33 and 83 are under examination. Claims 2-4, 6-8, 17-31 and 34-82 are withdrawn from consideration. Claims 16, 32-33 and 83 are objected to allegedly because these claims recite nonelected subject matter in the alternative. Claims 1, 5, 13, 32-33 and 83 are rejected under 35 U.S.C. § 112, first paragraph for allegedly not being enabled by the specification. Claims 1, 5, 13, 32-33 and 83 are rejected under 35 U.S.C. § 112, first paragraph, for allegedly not satisfying the written description requirement. The Examiner also states that Applicants' Information Disclosure Statement submitted on April 29, 2008 has been considered, but no art rejection is raised because the Examiner is unable to establish a relationship between instant SEQ ID NO: 7 and the sequence referred to by Mack et al. (GenBank AB033025).

This Response addresses each of the Examiner's rejections and objections. Applicants therefore respectfully submit that the present application is in condition for allowance. Favorable consideration of all pending claims is therefore respectfully requested.

Amendments to Claims

Independent claims 1 and 16 have been amended to further define the biological sample as a "blood, serum, stool or gastrointestinal tract sample". Support for this amendment is found in the specification, e.g., page 24, lines 14-26.

Independent claims 1 and 16 have also been amended to delete the term "substantially" in reference to sequence identifiers. Further, claim 1 has been amended to delete the expression "predisposition to the onset".

No new matter is introduced by the foregoing amendments.

Examiner's Remarks Regarding KIAA1199

Before responding to the Examiner's rejections and objections, Applicants wish to first address the Examiner's remarks with respect to KIAA1199. Specifically, the Examiner states that she is unable to establish a relationship between instant SEQ ID NO: 7 and the KIAA1199 sequence referred to by Mack et al. See page 2, Section 4 of the Office Action. On page 16 of the Action, the Examiner further state that there is no record in the prior art relating to KIAA1199 or Genbank Accession No. NC-000015. The Examiner indicates that she attempted to blast the KIAA1199 mRNA sequence against SEQ ID NO:7 and obtained no similarities.

Applicants respectfully submit that it is a matter of routine procedure to conduct a BLAST search using SEQ ID NO: 7 as the query sequence. The results of such a BLAST search are in fact extremely clean in that hits are only obtained with respect to KIAA1199, with the next closest hits exhibiting less than 4% query coverage. Applicants are providing herewith as **Exhibit 1**, a document which summarizes the BLAST results of SEQ ID NO: 7 and which demonstrates that the sequence which one obtains is the KIAA1199 sequence. SEQ ID NO: 7 aligns to the map region 51,882,643-51,915,746 on chromosome 15.

Objection to Claims

Claims 16, 32-33 and 83 are objected to allegedly because these claims recite nonelected subject matter in the alternative.

Applicants respectfully disagree with the Examiner. Independent claim 16 is directed to detection of co-expression of two or more nucleic acid molecules, at least one of which is the elected nucleic acid molecule comprising SEQ ID NO: 7. That is, the elected nucleic acid molecule which comprises SEQ ID NO: 7 is analyzed *together with* any one or more of the other

SEQ ID NOs. recited in subparagraph (i). Assuming that the use of SEQ ID NO:7 is ultimately held to be patentable, the method of claim 16 based on the combined use of SEQ ID NO: 7 with one or more other nucleic acids should also be found patentable. It is respectfully submitted that the claims, as presently recited, properly reflect the elected subject matter (SEQ ID NO: 7).

Accordingly, the objection to the claims is overcome and withdrawal thereof is respectfully requested.

35 U.S.C. § 112, First Paragraph – Enablement

Claims 1, 5, 13, 32-33 and 83 are rejected under 35 U.S.C. § 112, first paragraph, for allegedly not being enabled by the specification. With respect to the Examiner's analysis based on the *In re Wands* factors, Applicants respectfully submit the following and maintain that those skilled in the art would be able to practice the methods, as presently claimed, without undue experimentation.

Nature and Breadth of the Invention

The Examiner has noted that claim 1 continues to recite the language "or a predisposition to the onset" of an adenoma. This language has been deleted from claim 1 by way of the instant amendment. As amended, claim 1 is directed to a method of determining the onset of a colorectal adenoma.

The Examiner has also interpreted the language of "measuring the level of expression" of a nucleic acid molecule as including the measurement of transcription or translation of a nucleic acid molecule. Applicants submit that the Examiner's interpretation is correct in this regard. Clearly, methods for measuring transcription and translation of nucleic acid molecules are well known and have been widely used for many years.

With respect to the claim language referencing nucleic acid molecules hybridizing to SEQ ID NO:7 under high stringency conditions, the Examiner asserts that many nucleic acid molecules would hybridize to SEQ ID NO:7 under these conditions, including homologs or variant of SEQ ID NO: 7. However, the Examiner does not appear to have appreciated that high stringency conditions, by their very nature, permit hybridization of molecules exhibiting only very high levels of sequence identity with SEQ ID NO: 7. Accordingly, those skilled in the art would understand that the claims would not encompass a wide class of molecules unrelated to SEQ ID NO: 7. Further, under the law, Applicants are not required to limit the claims to those specifically exemplified embodiments; but rather, are permitted to have a breadth of claimed subject matter which is consistent with the disclosure of the specification. In re Anderson, 176 USPQ 331, 333 (CCPA 1973). Based on the guidance provided in the specification, those skilled in the art would be able to identify nucleic acid molecules capable of hybridizing to SEQ ID NO: 7 under high stringency conditions, and to further use the identified molecules in the claimed methods, without undue experimentation. Applicant's position in this regard is further supported by the discussion in relation to KIAA1199 below.

The Examiner continues to assert that the claims encompass within their scope "homologues, variants or the like" of SEQ ID NO: 7. However, this text was already deleted from the claims in response to the previous Office Action. As submitted above, while the claims do encompass related hybridizing molecules, such molecules must share a high level of identity with SEQ ID NO: 7.

With regard to the Examiner's rejection of the breadth of the term "biological sample", Applicants have amended the claims to define the sample as a blood, serum, stool or

gastrointestinal tract sample. Applicants respectfully submit that the specification specifically discloses that change of expression can be detected in a blood, serum, stool or gastrointestinal tract sample. See, e.g., page 24, lines 14-26 of the specification. Given that a change in expression is observed and has been specifically demonstrated in adenoma tissue (see, e.g., pages 100-107 of the specification), those skilled in the art would reasonably expect that the change would also be found in the stool, since the adenoma cells are shed into the stool. In addition, those skilled in the art would also reasonably expect that the change in expression could be detected in the blood and serum. As support of Applicant's position in this regard, it is submitted that it has been documented in the art that change in marker levels of colorectal neoplastic tissue could be detected in the blood, as reported by, e.g., Park, *Oncology* 22: 147 (2008) "Biomarker CCA-2 may provide accurate blood test for colorectal cancer", and Walgenbach-Brunagel et al., *J. Cell. Biochem.* 104: 286-294 (2008) (**Exhibit 2**). There also exist classic examples of serum biomarkers for other types of cancers, including the Prostate Cance Antigen (PSA), carcinoembryonic antigen (CEA), and carbohydrate antigen 19-9 (CA 19-9).

Guidance in the specification and working examples

The Examiner alleges on page 6 of the Office Action that "there is no external working example which validates the use of SEQ ID NO: 7 as a marker for colorectal adenoma." The Examiner apparently recognizes that the specification discloses that clones 8-2d and 12-2f, to which SEQ ID NO: 7 corresponds, were up-regulated by 50 and 45 fold, respectively, in adenoma tissue samples. However, the Examiner states that the specification is silent on how these two clones are actually related, or how these clones relate to SEQ ID NO: 7.

In the first instance, Applicants respectfully submit that 8-2d and 12-2f are two different clones which were generated from the same starting materials. Both clones express the SEQ ID NO: 7 sequence. It is not uncommon that more than one clone expressing the same nucleotide sequence was generated. Both clones were identified as corresponding to SEQ ID NO:7 and the results obtained using both these clones were consistent, with 8-2d showing a 50-fold increase relative to mean expression levels of normal tissue, and 12-2f showing an average of 45-fold increase relative to mean expression levels. Therefore, the Examiner's allegation that the specification does not validate the use of SEQ ID NO: 7 as a marker of colorectal adenoma is unfounded.

Applicants are providing herewith additional support for the use of SEQ ID NO: 7 as a marker of colorectal adenoma. As discussed above, SEQ ID NO: 7 corresponds to the KIAA1199 gene. **Exhibit 3** is an extract from a company report prepared in July of this year, which relates to KIAA1199 and demonstrates that it exhibits a 25-fold higher mean expression in adenoma than in normal tissue when tested across 19 patients and 30 normal individuals. Applicants provide the following additional information in relation to the data presented in **Exhibit 3**:

(i) The analyses which were performed looked at three regions of KIAA1199, of which one was the SEQ ID NO: 7 region. All three regions produced results which confirm that the expression of KIAA1199 was significantly increased in adenoma versus normal tissue. As can be seen from the results, there do occur a few isolated samples in the normal tissues which show a higher level of KIAA1199 expression as there occur a few adenoma tissue samples which show lower levels of KIAA1199. However, the mean results clearly indicate a very significant increase in KIAA1199 expression relative to normal tissue, fully supporting the claimed methods based on detecting an increased level of expression of a nucleic acid comprising SEQ ID NO:7 as "indicative" of the onset of an adenoma. These results in **Exhibit 3** are consistent with the results provided in the specification and

conclusively demonstrate the value of screening for increased expression of a nucleic acid molecule comprising SEQ ID NO: 7 (KIAA1199) as a marker of adenoma onset.

(ii) These results confirm that one can screen for any region of KIAA1199, being the gene comprising and characterized by SEQ ID NO: 7.

(iii) These results are also consistent with the results obtained and described in the specification using clones 8-2d and 12-2f, although not precisely identical in terms of overall fold increase. Those skilled in the art would appreciate that within any screening system there will always occur a certain standard of deviation where one performs the same experiment, with the same reagents, two or more times. However, both the results presented in **Exhibit 3** and the results described in the specification support the conclusion that increase in expression of a nucleic acid molecule comprising SEQ ID NO: 7 is indicative of adenoma.

The unpredictability of the art and the state of the art

The Examiner has again reiterated that the claimed diagnostic method relates to an extremely unpredictable art. In this regard, the Examiner has referred to results obtained in relation to prostate specific membrane antigen (PSMA). However, Applicants respectfully submit that the present application does not concern PSMA. Rather, it concerns a nucleic acid molecule comprising SEQ ID NO: 7. Based on the data provided in the specification and in the exhibits attached hereto, expression levels of a nucleic acid comprising SEQ ID NO: 7 have been shown to provide strong indication of adenoma development.

Further, Applicants respectfully submit that analysis of both nucleic acid expression and protein expression is widely used in many different diagnostic disciplines as a marker of the onset of certain disease conditions. The fact that such screening assays are so widely applied would suggest that the art is in fact not unpredictable and that PSMA, for example, may represent the

exception rather than the rule. To this end, the Examiner's assertion that it is highly unpredictable what level of expression of SEQ ID NO:7 must be observed in order to conclude that adenoma is present, particularly in light of the additional data provided in **Exhibit 3**, is arguably preposterous. The data clearly teach that a level of expression of adenoma above that of a normal level is indicative of adenoma. There is no need to focus on any specific expression level, rather, merely a level which is higher than the normal level, as would be understood by those skilled in the art. There is no extensive unpredictable experimentation which would be required to be undertaken in order to diagnose cancer. To this end, Applicants also attach an article by Sabates-Bellver (*Mol. Cancer Res.* 5: 1263-1275, 2007) (**Exhibit 4**), which was published subsequently to the filing of the present application and confirms that those in the art consider KIAA1199 as extremely important in the context of the diagnosis of adenoma.

Quantity of experimentation

The Examiner contends that it would require extensive experimentation before those skilled in the art could practice the claimed invention. In the first instance, the Examiner's attention is directed to the fact that reference to functional derivatives, variants, homologues and the like has been deleted from the claims. Although the claims encompass those that hybridize at levels of high stringency to SEQ ID NO:7, these hybridizing molecules must share a high level of identity with SEQ ID NO:7, as discussed above. One would expect that these hybridizing molecules would be molecules exhibiting minor variations in sequence, such as different isoforms of the molecule. To the extent that such isoforms may exist, the data presented in **Exhibit 3** certainly indicate that across 19 patients and 30 non-diseased control individuals, irrespective of the form of SEQ ID NO:7 which they may express, the levels of KIAA1199 are increased in

patients who have undergone the onset of an adenoma.

The Examiner further argues that it would require extensive experimentation to screen for a SEQ ID NO: 7 expression product. Applicants disagree with this assertion. Once a DNA molecule has been identified, its encoded expression product is routinely identifiable and can be analyzed by routine procedures. In this connection, once the recognition is provided by the present invention that the gene characterized by SEQ ID NO: 7 is diagnostic of the onset of adenoma, experimentation relating to establishing that SEQ ID NO: 7 characterizes KIAA1199, and experimentation relating to electing an appropriate form of KIAA1199 (e.g. DNA or protein) to screen for, or determining how screening assay should be conducted, are well within the scope of those skilled in the art, and would be a matter of routine procedure. In this connection, Applicants respectfully submit that additional experimentation is permissible. In re Wands, 858 F.2d 731, 736-737, 8 U.S.P.Q. 1400, 1404 (Fed Cir. 1988). Necessary experimentation is not determinative of the question of enablement; only undue experimentation is fatal under the provisions of 35 U.S.C. §112, first paragraph. Id. In the present case, Applicants submit that any additional experimentation, if needed, in order to practice the claimed invention is routine and not undue.

In sum, it is respectfully submitted that the evidence and guidance provided by the specification, together with the knowledge of persons of ordinary skill in the art, clearly enable those skilled in the art to practice the claimed methods, without undue experimentation. Therefore, it is respectfully requested that the Examiner withdraw the rejection based on the alleged non-enabling disclosure provided in the specification.

35 U.S.C. § 112, First Paragraph – Written Description

Claims 1, 5, 13, 32-33 and 83 are rejected under 35 U.S.C. § 112, first paragraph, for

allegedly not satisfying the written description requirement. The Examiner's rejection centers around the genus of nucleic acid molecules encompassed by the claims which are employed in the claimed methods.

Applicants first respectfully submit that the present claims do not include any reference to "functional derivative, variant or homologue". With respect to molecules that are capable of hybridizing to SEQ ID NO: 7 under high stringency conditions, as submitted, this is a relatively small and well defined genus, as nucleic acids that hybridize to SEQ ID NO: 7 under high stringency conditions must share a very high sequence identity with SEQ ID NO: 7.

Applicants further respectfully submit that nucleic acids claimed based on hybridization language may be considered to have met the written description requirement, because highly stringent hybridization conditions dictate that the species within the claimed genus are structurally similar, i.e., similar in sequence to the recited sequence in the claims. See, Enzo Biochem, In.c, v. Gen-Probe Inc., 323 F.3d 956, 967-968 (Fed. Cir. 2002).

Accordingly, Applicants respectfully submit that those skilled in the art would have considered that the specification adequately and clearly describes the genus of nucleic acids as presently claimed, and that the specification has conveyed to those skilled in the art that Applicants are in possession of the claimed genus. As such, it is respectfully requested that the written description rejection under 35 U.S.C. § 112, first paragraph be withdrawn.

Conclusion

In view of the foregoing amendments and remarks, it is firmly believed that the subject application is in condition for allowance, which action is earnestly solicited.

Respectfully submitted,

Xiaochun Zhu
Registration No. 56,311

SCULLY, SCOTT, MURPHY & PRESSER, P. C.
400 Garden City Plaza-STE 300
Garden City, New York 11530
(516) 742-4343
XZ:ab

Enc.: Exhibits 1-4

EXHIBIT 1

Executive Summary

The alignment search tool, BLAST, is available through NCBI, www.ncbi.nlm.nih.gov:

The "Nucleotide blast"[1] tool, BLASTn will search a NUCLEOTIDE database using a NUCLEOTIDE query.

SEQID7 (appendix 1) was queried using the standard settings in the NCBI/BLAST/blastn:

- 8 BLAST hits on the Query sequence
- Two significant hits were obtained Human Chromosome 15 genomic contig reference assembly NT_1094.16 and NW_001838219.1
- The remaining 4 hits had less than 4% query coverage.

The alignment of SEQID7 and the top hit, NT_010194.16 is given in Appendix 2

- Opening the link the matched reference sequence NT_010194.16 will result in a graphic demonstration of the relationship of SEQID7 to the KIAA1199 gene:

The KIAA1199 gene is in map region 51,862,032- 52,034,319 on Chr15
SEQID7 aligns to the map region 51,882,643 – 51,915,746

1 http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Nucleotides&PROGRAM=blastn&MEGABLAST=on&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on

APPENDIX 1 SEQID 7 NUCLEOTIDE SEQUENCE

aagcggctctcaaccgaaacctccccaggggctacagtggcctttccatgtggctttctcacagcatgttggctgtgttccaatggtgaaagt
ccacagagagagagagagacccagtggaaggcacatcatttttctaaacgactcctggaagttacacttctgctccatcctgggactacaggg
cacataacaccatgccagctaatttttatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgcgtagacgggatctcatcatcttgcc
caggttagtcttggactcctggactcaagtgatcctcccatctcaacctccaaaagtgctgggttacaggcatgagtcactgcacctggctgg
aaatttgttaatagcctatgttgaaggggtagctgaaatcacctcaccatcctctgggtttccagagcacctccattcttatagcccatgtga
gtcgtatggtggggtgtggttctatgtccttctcgccctctgtctgggactgccgagagagcaggtctcatgtcatttatttgtagagtctc
agggcctattgcaggatctgacagagtcaatgactgctacctttgcggaatgaatgaataaaatcattaatggctgaatgtggctggcttttc
cacgccttcccacagctggggtacttaatattggctgaggcaactacttttaaactgttggtatttctctttaataaaatcttgggaaaacct
tgactttcatgtcatttacttgggacttttccaaaatccaggctttatttttcatcaaacacatgtcatgatcatgctcttagggagtct
ttacaaaccatcatcatgctcttgagggaatcttttgaaaaccttactttagatcagagttagagaagaaattcacattctaatagatttgca
gggtaattgatattctcgccatctctgttcatatttgaaatatttcagtactcgatgtaggggcaaaaacattgagtttacaccttctaataa
ctttccaaaaacctgttataaagtaaaactgctgattcagaggtttggggatctctggggatacagctcagccttggggcccagggcctaccg
tagctgggctacaccttcctctccagcttcttgtccagctgcttctccctttctgttttagactctagcaacatcctaggattgttatggtcct
gttgatgcaatgctgcttcttgccatcttgctgctgtaaatgctgctttctctgctcaatcatctagcaaactaccattcattcttcctgacc
ctgctgaggcatccccttctctgtgaagagttccctctctccttctccaatgtatcagtaagctattgctgtgtaataaaccacccccaaggc
agtggcttgaaacaactgtgtattattgtcctgtgggtcaaccagctggttctgctgatccggacaggcttggctaatctcaactctgttttg
tatctatcggcagaacaactggaggctggctggtctaggatggcctcatttatgtgtttggcattggctagctctcaattcagtggatgaggg
tgactggaccatgcgtctctcatcacccagtagactagcctgggtttgttctcgaggtgactgatgctgttctgtgagagagaaggaaagcat
gcatggcctctggaggcctggatctcaaaacccaatggcacaccagcacttctgctgctttcttttggctgaagccaggtcagttcatgttca
aggggaggagatttagactctacctttaatgggagaagctgcatagttacattgcaaaagacaaggatctgaggggaggagagaaggatgag
tggaatggttgattgatttttgtgatcaatccaccacacccaccttgatagaggtacttactctgtagtacaattggccttccatactgtgg
gttccatacctatagattcaaccaattgcaaactgaaaatatttgaaatatgtttgcatctgcactgaacatgtacagactatttttcttgtc
cttacaggataataatacgggataacaactattgacaaagcatttacattgtattaggtattataagtaatctagagatgatttaaagtatac
aggaggatgtatgtatgttatatgcaaatactacactcttttatattagggacttgagcatctggagagtgtggtatctgagggagttcctgg
aactaatgtgcagatgccaagggacaactgtactattgtacttggaagtactcatgggtcatattgcattgtttctttgagtcctaattctg
ccaacatggcctggtgcttgcattaatcagctttctaatctctgagtaacaaggcacagtaacaaggagcagtaacaaggcacagggctggca
cctgagagtggaggtacccaggaggcagacaccataaggcgggaaagggacatatgtacagaatcatggctgcatgtcctgaagcctggctta
agccatcaacggctgctgggcagggccaaagccctgttatccctttcgccctcctgatggctctgccttcagctgggcgtgggca
ggccccacccaccgaggctccagccctacccacagtgtcagcaatgcagcctccagaggatgtgctcaggccctgcccacacacccggatgt
tgacaggggcatgactccagcgccagctctaatggatggtctcatcgcttttaaaataatgaccatgggcgtgggctggcgagagcagtgac
atcactttcctgcaattctgggtcagttcctgctgcttttctctgtatgtttgaatgactgaaataaatctattggttggatatatttcctgg
aagacttctgacatgttcacatgcctatcttggaatgtggtcaggagagcaatggctttggacttagaggtcctgggttcaagattctgctac
tagcttgctgtatgaacttggactagcaacttaacttctccaggcctgtgttttctcatttgtacaatgatgggaggaataccttggttttg
taaaggaatggtgaggacgaactgggatctcttgtcagagacactgtcttmgtcagtttgggctgctataacaaagttccacagattaggtgg
cttgttaacagcagaaatatatttctcacagttctggaggctagaagtccaagctcaggatgccagcatggttgggctctggagggctcctaa
acaccattattcttcattcacgcttctcagagccctaaggaagagagtgattcctcagctcaattgtgaactgctcctgccactctgtacttc
ctcgtgtaaagaarccagactttacatcatgggtgaccactcccgcagagttgtacagaacctcccttggggccacaggatggctggattctg
tcccctcatatacaaggaggttattgggacagcatttctccctagaacaagagtgtatatttcagaaagctatggatgacttcccatggtcat
cagatcactaggcaggaatgctattctcctgatagatgtgtggaaagtattcaattcaattttgacccaaagttctaggcactggattaagaa
atgccaaacccaaaacgtttaactttagaattaaaaaaaaaaaa

APPENDIX 2 BLASTn QUERY RESULT of SEQID7

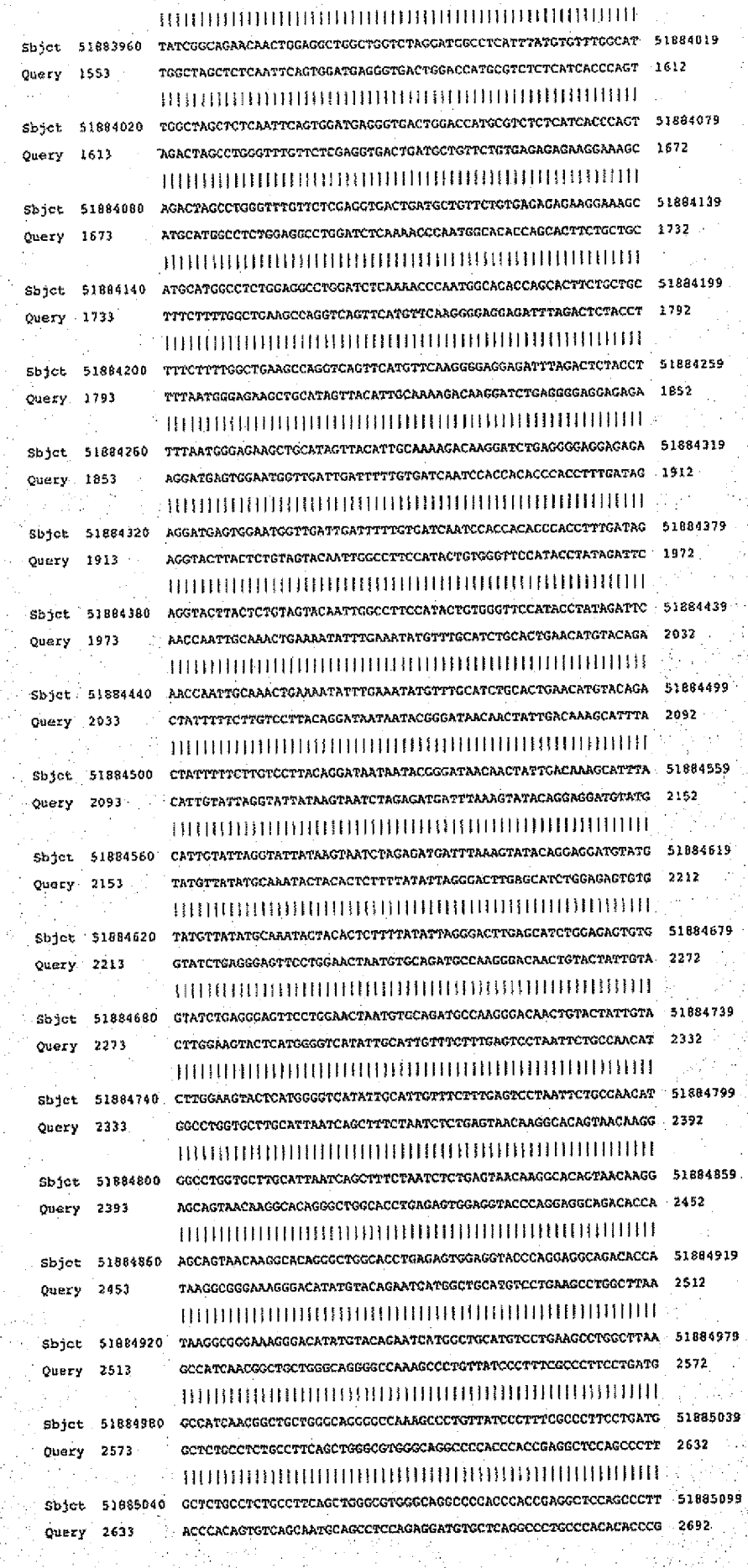
>ref|NT_010194.16|Hs15_10351 Homo sapiens chromosome 15 genomic contig, reference assembly
Length=53619965

Features flanking this part of subject sequence:
45612 bp at 5' side: hypothetical protein LOC58489
70405 bp at 3' side: KIAA1199

Score = 6408 bits (3470), Expect = 0.0
Identities = 3488/3497 (99%), Gaps = 3/3497 (0%)
Strand=Plus/Plus

```
Query  173       CTGGGACTACAGGGCACATAACACCATGCCAGCTAATTTTATGTGTGTGTGTGTGTGT  232
                  ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51882643  CTGGGACTACA-GGCACATAACACCATGCCAGCTAATTTTATGTGTGTGTGTGTGTGT  51882701
Query  233       GTGTGTGTGTGTGTGTGTGCGTAGACGGGATCTCATCATCTTGCCCAGGTTAGTCTTG  292
                  |||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Sbjct  51882702  GTGTGTGTGTGTGTGTGT--GTAGACGGGATCTCATCATCTTGCCCAGGTTAGTCTTG  51882759
Query  293       GACTCCTGGACTCAAGTGATCCTCCCATCTCAACCTCCAAAAGTGCTGGGTTACAGGCAT  352
                   ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51882760  AACTCCTGGACTCAAGTGATCCTCCCATCTCAACCTCCAAAAGTGCTGGGTTACAGGCAT  51882819
Query  353       GAGTCACTGCACCTGGCTGGAAATTTGTTAATAGCCTATGTTGAAGGGGTAGCTGAAATC  412
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Sbjct  51882820  GAGTCACTGCACCTGGCTGGAAATTTGTTAATAGCCTATGTTGAAGGGGTAGCTGAAATC  51882879
Query  413       ACCTCACCATCCTCTGGGTTTCCAGAGCACCTCCATTCTTATAGCCCATGTGAGTCGTAT  472

Sbjct  51882880  ACCTCACCATCCTCTGGGTTTCCAGAGCACCTCCATTCTTATAGCCCATGTGAGTCGTAT  51882939
Query  473       GGTGGGGTGTGGTTCTATGTCCTTTCTCGCCCTCTGTCTGGGACTGCCGAGAGAGCAGGT  532

Sbjct  51882940  GGTGGGGTGTGGTTCTATGTCCTTTCTCGCCCTCTGTCTGGGACTGCCGAGAGAGCAGGT  51882999
Query  533       CTCATGTCATTTATTTGTAGAGTCTCAGGGCCTATTGCAGGATCTGACAGAGTCAATGAC  592

Sbjct  51883000  CTCATGTCATTTATTTGTAGAGTCTCAGGGCCTATTGCAGGATCTGACAGAGTCAATGAC  51883059
Query  593       TGCTACCTTTGCGGAATGAATGAATAAAATCATTAATGGCTGAATGTGGCTGGCTTTTCC  652

Sbjct  51883060  TGCTACCTTTGCGGAATGAATGAATAAAATCATTAATGGCTGAATGTGGCTGGCTTTTCC  51883119
Query  653       ACGCCTTCCCACAGCTGGGGTACTTAATATTGGCTGAGGCAACTACTTTTAAACTGTTGG  712

Sbjct  51883120  ACGTCTTCCCACAGCTGGGGTACTTAATATTGGCTGAGGCAACTACTTTTAAACTGTTGG  51883179
Query  713       TATTTCTCTTTAATAAAATCTTGGGAAAACCTTGACTTTCATGTCATTTTACTTTGGGAC  772

Sbjct  51883180  TATTTCTCTTTAATAAAATCTTGGGAAAACCTTGATTTTCATGTCATTTTACTTTGGGAC  51883239
Query  773       TTTTTCCAAAATCCAGGCTTTATTTTTCATCAAACACATGTCATGATCATGCTCTTAGGG  832

Sbjct  51883240  TTTTTCCAAAATCCAGGCTTTATTTTTCATCAAACACATGTCATGATCATGCTCTTAGGG  51883299
Query  833       AGTCTTTACAAACCATCATCATGCTCTTGAGGGAATCTTTTGAAAACCTTACTTTAGATC  892

Sbjct  51883300  AGTCTTTACAAACCATCATCATGCTCTTGAGGGAATCTTTTGAAAACCTTACTTTAGATC  51883359
Query  893       AGAGTTAGAGAAGAAATTCACATTCTAATAGATTTGCAGGGTAATTGATATTCTCGCCAT  952

Sbjct  51883360  AGAGTTAGAGAAGAAATTCACATTCTAATAGATTTGCAGGGTAATTGATATTCTCGCCAT  51883419
Query  953       CTCTGTTCATATTTGAAATATTTCAGTACTCGATGTAGGGGCAAAAACATTGAGTTTACA  1012

Sbjct  51883420  CTCTGTTCATATTTGAAATATTTCAGTACTCGATGTAGGGGCAAAAACATTGAGTTTACA  51883479
Query  1013      CCTTCTAATAACTTTCCAAAAACCTGTTATAAAGTAAAACTGCTGATTCAGAGGTTTGGG  1072

Sbjct  51883480  CCTTCTAATAACTTTCCAAAAACCTGTTATAAAGTAAAACTGTTGATTCAGAGGTTTGGG  51883539

Query  1073      GATCTCTGGGGATACAGCTCAGCCTTGGGGCCCAGGGCCTACCGTAGCTGGGCTACACCT  1132

Sbjct  51883540  GATCTCTGGGGATACAGCTCAGCCTTGGGGCCCAGGGCCTACCGTAGCTGGGCTACACCT  51883599
Query  1133      TCCTCTCCAGCTTCTTGTCCAGCTGCTTCTCCCTTCTGTTTTAGACTCTAGCAACATCCT  1192

Sbjct  51883600  TCCTCTCCAGCTTCTTGTCCAGCTGCTTCTCCCTTCTGTTTTAGACTCTAGCAACATCCT  51883659
Query  1193      AGGATTGTTATGGTCCTGTTGATGCAATGCTGCTTCTTGCCATCTTGCTGCTGTAAATGC  1252

Sbjct  51883660  AGGATTGTTATGGTCCTGTTGATGCAATGCTGCTTCTTGCCATCTTGCTGCTGTAAATGC  51883719
Query  1253      TGCTTTCTCTGCTCAATCATCTAGCAAACTACCATTCATTCTTCCTGACCCTGCTGAGGC  1312

Sbjct  51883720  TGCTTTCTCTGCTCAATCATCTAGCAAACTACCATTCATTCTTCCTGACCCTGCTGAGGC  51883779
Query  1313      ATCCCCTTCTCTGTGAAGAGTTCCCTCTCTCCTTCTCCAATGTATCAGTAAGCTATTGCT  1372

Sbjct  51883780  ATCCCCTTCTCTGTGAAGAGTTCCCTCTCTCCTTCTCCAATGTATCAGTAAGCTATTGCT  51883839
Query  1373      GTGTAATAAACCACCCCCAAGGCAGTGGCTTGAAACAACTGTGTATTATTGTCCTGTGGG  1432

Sbjct  51883840  GTGTAATAAACCACCCCCAAGGCAGTGGCTTGAAACAACTGTGTATTATTGTCCTGTGGG  51883899
Query  1433      TCAACCAGCTGGTTCTGCTGATCCGGACAGGCTTGGCTAATCTCAACTCTGTTTTGTATC  1492

Sbjct  51883900  TCAACCAGCTGGTTCTGCTGATCCGGACAGGCTTGGCTAATCTCAACTCTGTTTTGTATC  51883959
Query  1493      TATCGGCAGAACAACTGGAGGCTGGCTGGTCTAGGATGGCCTCATTTATGTGTTTGGCAT  1552
```



```
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885100   ACCCACAGTGTCAGCAATGCAGCCTCCAGAGGATGTGCTCAGGCCCTGCCCACACACCCG  51885159

Query  2693       GATGTTGACAGGGGCATGACTCCAGCGCCAGCTCTAATGGATGGTCTCATCGCTTTTAAA  2752
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885160   GATGTTGACAGGGGCATGACTCCAGCGCCAGCTCTAATGGATGGTCTCATCGCTTTTAAA  51885219

Query  2753       ATAATGACCATGGGGCGTGGGCTGGCGAGAGCAGTGACATCACTTTCCTGCAATTCTGGG  2812
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885220   ATAATGACCATGGGGCGTGGGCTGGCGAGAGCAGTGACATCACTTTCCTGCAATTCTGGG  51885279

Query  2813       TCAGTTCCTGCTGCTTTTCTCTGTATGTTTGAATGACTGAAATAAATCTATTGGTTGGAT  2872
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885280   TCAGTTCCTGCTGCTTTTCTCTGTATGTTTGAATGACTGAAATAAATCTATTGGTTGGAT  51885339

Query  2873       ATATTTCCTGGAAGACTTCTGACATGTTCACATGCCTATCTTGGAATGTGGTCAGGAGAG  2932
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885340   ATATTTCCTGGAAGACTTCTGACATGTTCACATGCCTATCTTGGAATGTGGTCAGGAGAG  51885399

Query  2933       CAATGGCTTTGGACTTAGAGGTCCTGGGTTCAAGATTCTGCTACTAGCTTGCTGTATGAA  2992
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885400   CAATGGCTTTGGACTTAGAGGTCCTGGGTTCAAGATTCTGCTACTAGCTTGCTGTATGAA  51885459

Query  2993       CTTGGACTAGCAACTTAACTTCTCCAGGCCTGTGTTTTCTCATTTGTACAATGATGGGAG  3052
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885460   CTTGGACTAGCAACTTAACTTCTCCAGGCCTGTGTTTTCTCATTTGTACAATGATGGGAG  51885519

Query  3053       GAATACCCTTGGTTTTGTAAAGGAATGGTGAGGACGAACTGGGATCTCTTGTCAGAGACA  3112
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885520   GAATACCCTTGGTTTTGTAAAGGAATGGTGAGGACGAACTGGGATCTCTTGTCAGAGACA  51885579

Query  3113       CTGTCTTMGTCAGTTTGGGCTGCTATAACAAAGTTCCACAGATTAGGTGGCTTGTTAACA  3172
                  ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885580   CTGTCTTAGTCAGTTTGGGCTGCTATAACAAAGTTCCACAGATTAGGTGGCTTGTTAACA  51885639

Query  3173       GCAGAAATATATTTCTCACAGTTCTGGAGGCTAGAAGTCCAAGCTCAGGATGCCAGCATG  3232
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885640   GCAGAAATATATTTCTCACAGTTCTGGAGGCTAGAAGTCCAAGCTCAGGATGCCAGCATG  51885699

Query  3233       GTTGGGCTCTGGAGGGCTCCTAAACACCATTATTCTTCATTCACGCTTCTCAGAGCCCTA  3292
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885700   GTTGGGCTCTGGAGGGCTCCTAAACACCATTATTCTTCATTCACGCTTCTCAGAGCCCTA  51885759

Query  3293       AGGAAGAGAGTGATTCCTCAGCTCAATTGTGAACTGCTCCTGCCACTCTGTACTTCCTCG  3352
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885760   AGGAAGAGAGTGATTCCTCAGCTCAATTGTGAACTGCTCCTGCCACTCTGTACTTCCTCG  51885819

Query  3353       TGTAAAGAARCCAGACTTTACATCATGGGTGACCACTCCCGCAGAGTTGTACAGAACCTC  3412
                  ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885820   TGTAAAGAAACCAGACTTTACATCATGGGTGACCACTCCCGCAGAGTTGTACAGAACCTC  51885879

Query  3413       CCTTGGGGCCACAGGATGGCTGGATTCTGTCCCCTCATATACAAGGAGGTTATTGGGACA  3472
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885880   CCTTGGGGCCACAGGATGGCTGGATTCTGTCCCCTCATATACAAGGAGGTTATTGGGACA  51885939

Query  3473       GCATTTCTCCCTAGAACAAGAGTGTATATTTCAGAAAGCTATGGATGACTTCCCATGGTC  3532
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51885940   GCATTTCTCCCTAGAACAAGAGTGTATATTTCAGAAAGCTATGGATGACTTCCCATGGTC  51885999

Query  3533       ATCAGATCACTAGGCAGGAATGCTATTCTCCTGATAGATGTGTGGAAAGTATTCAATTCA  3592
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51886000   ATCAGATCACTAGGCAGGAATGCTATTCTCCTGATAGATGTGTGGAAAGTATTCAATTCA  51886059

Query  3593       ATTTTGACCCAAAGTTCTAGGCACTGGATTAAGAAATGCCAAACCCAAAACGTTTAACTT  3652
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  51886060   ATTTTGACCCAAAGTTCTAGGCACTGGATTAAGAAATGCCAAACCCAAAACGTTTAACTT  51886119

Query  3653       TAGAATTAAAAAAAAAA  3669
                  |||||||||||||||||
Sbjct  51886120   TAGAATTAAAAAAAAAA  51886136
```

Features flanking this part of subject sequence:

78543 bp at 5' side: hypothetical protein LOC58489

40295 bp at 3' side: KIAA1189

Score = 296 bits (160), Expect = 1e-76

Identities = 169/173 (97%), Gaps = 1/173 (0%)

Strand=Plus/Plus

```
Query  5         GGCTCTCAACCGAAACCTCCCCAGGGGCTACAGTGGCCTTTCCATGTGGCTTTCTCACAG  64
                 ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct  51915574  GGCTCTCAACCGAAACCTCCCCAGGGGCTACAGTGGCCTTTCCGTGTGGCTTTCTCACAG  51915633
Query  65        CATGTTGGCTGTGTTCCAATGGTGAAAGTCCACAGAGAGAGAGAGAGACCC-AGTGGAAG  123
                 ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Sbjct  51915634  CATGTTGGCTGTGTTCCAATGGTGAAAGTCCACAGAGAGAGAGAGAGACCCCAGTGGAAG  51915693
Query  124       GCACATCATTTTTCTAAACGACTCCTGGAAGTTACACTTCTGCTCCATCCTGG  176
                 ||||||||||||||||||||||| |||||||||||||||||||||||| ||||
Sbjct  51915694  GCACATCATTTTTCTAAACGACTCTTGGAAGTTACACTTCTGCTCCATTCTGG  51915743
```

EXHIBIT 2

Journal of Cellular Biochemistry 104:286–294 (2008)

The Use of a Colon Cancer Associated Nuclear Antigen CCSA-2 for the Blood Based Detection of Colon Cancer

Gisela Walgenbach-Brunagel,[2] Bettina Burger,[2] Eddy S. Leman,[1] Klaus-J. Walgenbach,[2] Rene Tolba,[3] Lukas Heukamp,[4] Andreas Hirner,[2] and Robert H. Getzenberg[1]*

[1]Department of Urology, Johns Hopkins University School of Medicine, Baltimore, Maryland
[2]Department of Surgery, University of Bonn, Bonn, Germany
[3]House of Experimental Therapy, University of Bonn, Bonn, Germany
[4]Department of Pathology, University of Bonn, Bonn, Germany

Abstract The early diagnosis of colorectal cancer (CRC) is central for effective treatment, as prognosis is directly related to the stage of the disease. Development of tumor markers found in the blood from patients, which can detect CRC at an early stage, should have a major impact in morbidity and mortality of this disease. The nuclear matrix is the structural scaffolding of the nucleus and specific nuclear matrix proteins (NMPs) have been identified as an "fingerprint" for various cancer types. Previous studies from our laboratory have identified four colon cancer associated NMPs termed colon cancer-specific antigen (CCSA)-2 to (CCSA)-5. The objective of the present study was to analyze the expression of one of these proteins, CCSA-2 in serum from various patient populations and to determine whether CCSA-2 antibodies could be used in a clinically applicable serum-based immunoassay specifically to detect colon cancer. Using an indirect ELISA, which detects CCSA-2, the protein was measured in the serum from 174 individuals, including healthy individuals, patients with colon cancer, patients with diverticulosis, colon polyps, inflammatory bowel disease (IBD) as well as other cancer types. With a predetermined cutoff absorbance of 0.6 OD we have successfully utilized this approach to develop an immunoassay that detected colon cancer. The immunoassay showed a sensitivity of 88.8% (24/27) and an overall specificity of 84.2% (106/127). This initial study showed the potential of CCSA-2 to serve as a highly specific blood based marker for colon cancer. Although potentially promising, the results of this study must be confirmed in larger independent validation studies. J. Cell. Biochem. 104: 286–294, 2008.

Key words: nuclear matrix proteins; colorectal cancer; tumor markers

Colorectal cancer (CRC) is one of the best characterized tumor types in regards to the multistep genetic progression pathway that has been elucidated. Despite our molecular understanding it is the second leading cause of cancer related death in the United States and the third most common cancer after lung and breast cancer worldwide [Parkin, 2001].

In 2007, more than 153,760 new cases will be diagnosed and more than 52,180 people will die from CRC in the USA [Jemal et al., 2007]. More than 50% of these deaths may have been prevented through the use of screening tests as the resulting early detection of the disease [Walsh and Terdiman, 2003]. The long natural history of CRC as it evolves from adenomatous polyps in the majority of cases provides opportunities for detection of early stage in cancer and for prevention of cancer by removal of polyps. Despite the potential for screening of CRC, only a minority of the population currently undergo screening program (www.cancer.org).

Dr. R.H. Getzenberg and Dr. G. Walgenbach-Brunagel hold a patent for the technology described in this article. This patent is owned by the University of Pittsburgh and has been licensed to Onconome, Inc. Dr. R.H. Getzenberg is a consultant to the company. None of the other authors have relationships related to this work.

Grant sponsors: Else Kröner-Fresenius-Stiftung; BONFOR Programme of the Medical Faculty, University of Bonn; Ministerial of Nordrhein-Westfalen, Germany.

*Correspondence to: Robert H. Getzenberg, PhD, Director of Research and Professor, Johns Hopkins University School of Medicine, Department of Urology, Marburg 121, 600 N Wolfe Street, Baltimore, MD 21287-2101.
E-mail: rgetzen1@jhmi.edu

Received 18 September 2007; Accepted 26 September 2007
DOI 10.1002/jcb.21619

The low rate of participation in CRC screening is critical to understand and is due to a number of actors, including patient discomfort, costs, and poor acceptability of current screening methods. Compliance to a serum test likely be better than tests involving feces and stool handling. An effective blood test, which ideally has a high specificity and sensitivity would be an ideal method to detect CRC and could lead to a reduction of the mortality and morbidity of CRC.

In order to identify highly specific tumor markers, investigators have focused attention on the structural changes that are associated with neoplastic transformation. Alterations in the cellular and nuclear structure are hallmarks of the carcinogenic process. These alterations are so prevalent in cancer cells that they are commonly used as a pathological marker of transformation. Nuclear shape reflects the internal nuclear structure and processes and is determined by the nuclear matrix [Pienta et al., 1989].

Most of the nuclear matrix proteins (NMPs) identified to date are common to all cell types, but several identified NMPs are tissue and cell line specific [Getzenberg, 1994]. This structure has many important functions like DNA organization, stabilization, and organization of gene regulatory complexes and synthesis of RNA, a variety of functions of which many have implications in cancer progression [Konety and Getzenberg, 1999].

Cell type-specific "fingerprinting" of aberrant NMPs and their appearance in cancer development has led to the analysis of NMP composition of a variety of tumors in an effort to determine whether these proteins can be developed as diagnostic and/or prognostic markers for cancer. Previously, we have identified specific NMP in prostate, bladder, renal, colon cancer, and colon cancer metastasis to the liver [Konety et al., 1998; Brunagel et al., 2004, 2002a,b; Van Le et al., 2004; Myers-Irvin et al., 2005; Paul et al., 2005]. This oncological "fingerprint" can be used as a specific and reliable diagnostic test, even when a distinction may not be made accurately on a histological basis alone [Getzenberg et al., 1991; Dhir et al., 2004a].

Our laboratory has recently demonstrated, that an antibody raised against the prostate cancer associated marker EPCA-2 is a sensitive and specific serum test for prostate cancer [Dhir et al., 2004b; Paul et al., 2005; Leman et al., 2007]. Additionally, an enzyme linked immunosorbent assay (ELISA) has been developed to detect a specific nuclear protein, BLCA-4 in the urine of individuals with bladder cancer. The test has shown to have a 96.4% sensitivity and 100% specificity [Konety et al., 2000].

Our previous studies describe the isolation of four NMPs (CCSA-2–CCSA-5) that are specifically expressed in colon cancer [Brunagel et al., 2002b]. One of these proteins, CCSA-2 was isolated by excising gel spots from negatively stained two-dimensional gels. The gels spots were then concentrated to obtain protein sequences and synthesized for antibody production.

Internal peptide sequencing of CCSA-2 resulted in four distinct peptides with sufficient amino acid sequence data. The four peptides along with the most significant matches obtained from BLAST analysis are described previously [Brunagel et al., 2002b]. Overall, while these data suggest that some regions of CCSA-2 may be common to other proteins, there is a high possibility of it being a novel uncharacterized protein.

The development of antibodies identifying aberrant NMPs in CRC could become clinically important assay with great specificity. The objective of this study was to investigate whether the NMP CCSA-2 can function as a highly specific and sensitive serum based biomarker for CRC.

Using an indirect ELISA approach, sera from patients with colon cancer were compared with serum samples from healthy donors, patients with diverticulosis/diverticulitis, patients with inflammatory bowel disease (IBD), patients with colon polyps, patients after curative treatment of colon cancer and patients with different cancers.

MATERIALS AND METHODS

Protein Sequencing

CCSA-2 was isolated according to an adaptation of a technique developed by Gevaert [1995]. Two-dimensional gels were negatively stained by 0.2 M imidazole and 0.3 M zinc chloride. The staining was stopped, and the protein gel spots were excised and frozen at −80°C. The spots were then stained with Coomassie blue and concentrated on an acrylamide/agarose gel and sequenced (Michigan State University).

Antibody Production

A standard protocol was followed in the production of monospecific antibodies raised against the CCSA-2 peptides in rabbits. Peptide sequences were chosen based upon the length of the sequence obtained as well as antigenicity. The peptide sequences were modified to contain a terminal cysteine for coupling purposes and conjugated to keyhole limpet hemocyanin or bovine serum albumin to increase immunoreactivity. Antibodies were produced at Biogenes Berlin (Germany) under an Institutional Animal Care and Use Committee approved protocol.

Patients

Serum samples were obtained from consenting patients under an Institutional Review Board—approved protocol. Serum samples from 174 patients were analyzed. After obtaining a blood sample, patients underwent a colonoscopy. Blood was collected with the blood collection system S-Monovette (Sarstedt, Nümbrecht, Germany). After collection, samples were centrifuge at 4,000 rpm. The supernatant was aliquoted in 2 ml tubes (Greiner Bio-one, Solingen, Germany,). The samples were stored at −80°C according to GLP (Good Laboratory Practice) conditions.

Of the patients studied, 27 were diagnosed with colon cancer. The control group consisted of 40 patients with a normal colon as evident by colonoscopy, 21 patients with a diverticulosis, 20 patients with colon polyps, 11 patients with an IBD, and 37 patients with different cancer types. Additionally nine patients 2–9 years after curative surgery for colon cancer were analyzed. The patient's characteristics are summarized in Table I.

Indirect ELISA

The detectability of CCSA-2 using the anti CCSA-2-antibody was assessed using serial dilutions of BSA-conjugated anti CCSA-2 antiserum against known concentrations of CCSA-2 peptide coated into a 96-well plate.

Using Nunc Immunoplate Maxisorb plates prepared with 50 µl coating solution (KPL, Baltimore, MD), 50 µl of serum per well, in triplicate, was allowed to incubated at room temperature with moderate shaking overnight. As a positive control, 50 µl of unlabeled rabbit immunoglobulin G (IgG), diluted with 50 µl coating solution (KPL), was plated overnight as well. The following day, all wells except the blank wells were blocked with 250 µl of Super Block Blocking Buffer (TBS; Pierce, Rockford, IL) for 45 min at 37°C. After blocking the wells, all wells were washed 3× with 250 µl reagent quality water before the addition of the primary antibody. The primary antibody for the sample wells consisted of 100 µl of diluted polyclonal antibody (previously described) in Super Block Blocking Buffer (Pierce). The negative control wells contained rabbit preimmune serum. Following a 2-h incubation period at 37°C with moderate shaking, the plate was emptied, washed with reagent quality water (250 µl, 3×), and then secondary antibody was added to all the wells for another 2 h. The secondary antibody applied was 1 mg/ml goat anti-rabbit IgG-horseradish peroxidase (human serum adsorbed) (KPL), diluted 1:5,000 in Super Block Blocking Buffer (Pierce). After washing the wells with reagent quality water (3 × 250 µl), 100 µl of 3,3′,5,5′-tetramethylbenzidine (KPL), was added to each well and allowed to react for 14 min and the absorbance was read at 650 nm on a Safire (Tecan, Germany) micro plate reader.

Statistical Analysis

The data were compiled as mean ± standard error of the mean.

The normal distribution of the samples of each group was controlled by the Kolmogorov Smirnov test. To analyze differences between the groups, one-way analysis of variance (ANOVA) with the Dunnett's post hoc test was performed. The colon cancer group was taken as reference, statistical significance was assumed at $P < 0.05$.

All statistical analysis and receiver-operator characteristic (ROC) curve were performed using GrapPad Prism version 4.03 for Windows XP, GraphPad Software (San Diego, CA, www.graphpad.com).

RESULTS

Using anti-CC2 antibodies, an indirect ELISA was developed to measure the level of CCSA-2 in the serum from various patient populations. The average value for CCSA-2 in the serum of the 27 colon cancer patients was 0.73 ± 0.15 OD, whereas the average value for healthy individuals (control) was 0.53 ± 0.06 OD. Statistical analysis demonstrated a highly significant difference in serum CCSA-2 levels between the

TABLE I. Characteristics of Patients Studied

	Colon cancer, n = 27	Control, n = 40											
A: Tumor stage (UICC) and grade of the 27 colon cancer patients and control (normal colonoscopy) n = 40													
Female	13	23											
Male	14	17											
Age (mean)	66.3 years	59 years											
Age (range)	39–78 years	27–70 years											
Tumor stage UICC													
Tis (carcinoma in situ)	1												
I: $T_{1-2}N_0M_0$	8												
II: $T_{3-4}N_0M_0$	10												
III: $T_{1-4}N_{1-2}M_0$	9												
IV: $T_{1-4}N_{1-2}M_1$	4												
Tumor grade													
G1	1												
G2	12												
G3	4												
	Diverticulosis/ diverticulitis n = 14	Inflammatory bowel disease, n = 11 M. Crohn, n = 7; C. ulcerosa, n = 4	Colon polyps, n = 20	Other inflammatory disease, n = 9 Gastritis, cholecysto-lithiasis pancreatitis	Cholangioca., n = 6	Pancreatic cancer, n = 11	HCC, n = 3	Lung cancer, n = 4	After colon cancer, n = 9 2–9 years after curative surgery	gastric cancer, n = 13
B: All other patients profile										
Female	10	4	4	4	1	5	2	1	7	4
Male	4	3	16	5	5	6	1	3	2	9
Age (mean) years	64	59.6	69.6	62.7	57.7	47	64	65.9	69	66.9
Age (range) years	53–74	41–76	35–83	31–69	54–67	23–67	40–71	57–70	46–78	45–78

CC, cholangio carcinoma; HCC, hepatocellular carcinoma.

TABLE IIa. Dunnett multiple Comparison test

Population pairs	Significance (P)
Colon cancer group taken as reference	
Colon cancer vs. control	$P < 0.01$
Colon cancer vs. diverticulosis/itis	$P < 0.01$
Colon cancer vs. IBD	$P < 0.01$
Colon cancer vs. colon polyps	$P < 0.01$
Colon cancer vs. various cancer types	$P < 0.01$
Colon cancer vs. healthy patients after colon cancer	$P < 0.01$
Colon cancer vs. inflammatory disease	$P < 0.01$

colon cancer patients and each of the other patients groups (Tables IIa and IIb).

The receiver operating characteristic curves for CCSA-2 are shown in Figure 1A,B. The CCSA-2 assay was highly accurate in separating colon cancer from healthy control (area under the curve 0.94, 95% confidence interval, CI, 0.89–0.99; Table III). Additionally the ROC curve was highly accurate in separating colon cancer from healthy control and all other patients population (area under the curve 0.8938 95% CI, 0.83–0.94). Using the ROC curve the cut off level of 0.6 OD was selected (Fig. 2).

Using a cutoff value from 0.6 OD the sensitivity was 88.8%; 24 from 27 colon cancer patients are detectable in serum with CCSA-2 and the specificity was 92.5%, 37 healthy individuals from 40 were identified with the assay as correctly negative. The overall specificity was 84.2%, 106 of the 127 individuals were diagnoses as normal were below the cut off (Table IV; Fig. 3).

DISCUSSION

The early diagnosis of CRC and the early detection of recurrence are central to the effective treatment of this disease. There is a consensus that CRC screening is effective and it can be prevented in many cases. Due to CRC screening the incidence of CRC has dropped in recent years, possibly due to the screening program [Mandel, 2005]. There is less consensus regarding optimal screening strategies, as sensitivity and specificity, and patient acceptance, limit current options. To overcome these barriers a range of approaches, including proteomics based testing, stool genetic testing, radiological imaging, and enhanced endoscopies has been the focus of intense research.

Presently, colonoscopy with a sensitivity of 97% and a specificity of 98% and a sensitivity of adenomas of at least 1 cm diameter of around 90% [Pickhardt et al., 2003; Winawer et al., 2003], is considered the gold standard for colon cancer diagnosis and offers the potential to both, find and remove premalignant lesions, but it is associated with high cost, patient discomfort, complication, and variable sensitivity given through the experience of the endoscopies.

A useful diagnostic assay must be sensitive and must detect the cancer in an early tumor

TABLE IIb. Summary of Data

Group	N	Mean	Standard deviation	Standard error of mean	Median
Colon cancer	27	>0.7359	>0.1576	>0.03034	>0.6900
Control	40	>0.5324	>0.06086	>0.009624	>0.5300
IBD	11	>0.5390	>0.06725	>0.02028	>0.5660
Colon polyps	20	>0.5686	>0.1329	>0.02971	>0.5400
Diverticulosis/itis	21	>0.5529	>0.09961	>0.02174	>0.5500
Other cancer type	37	>0.5687	>0.1445	>0.02375	>0.5450
After colon cancer	9	>0.4882	>0.05895	>0.01965	>0.4860
Inflammatory disease	9	>0.5573	>0.09594	>0.03198	>0.5340

Group	Minimum	Maximum	95% Confidence interval	
			From	To
Colon cancer	>0.5480	>1.250	>0.6735	0.7983
Control	>0.3990	>0.7200	>0.5129	0.5518
IBD	>0.4390	>0.6470	>0.4938	0.5841
Colon polyps	>0.3400	>0.8480	>0.5064	0.6308
Diverticulosis/itis	>0.4000	>0.7700	>0.5075	0.5982
Other cancer type	>0.3688	>1.031	>0.5204	0.6169
After colon cancer	>0.3920	>0.5690	>0.4429	0.5335
Inflammatory disease	>0.4390	>0.7300	>0.4836	0.6311

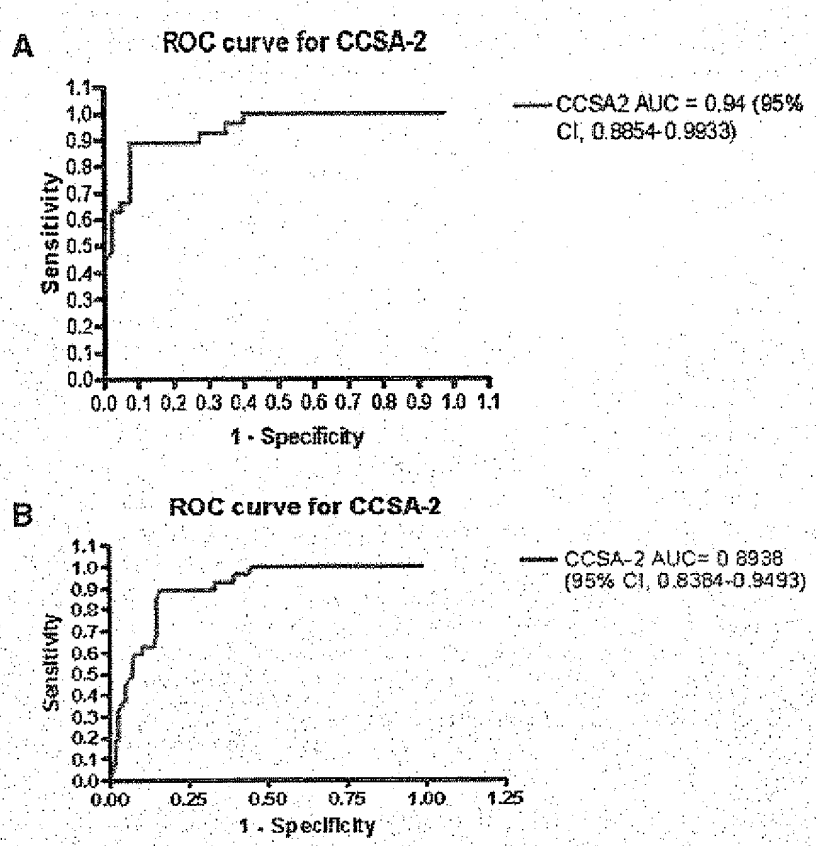


Fig. 1. A: Receiver-operator characteristic (ROC) curve for CCSA-2 in separating normal healthy patients and colon cancer patients. AUC: area under the ROC curve. B: Receiver-operator characteristic (ROC) curve for CCSA-2 in separating colon cancer patients from all other patients including healthy controls.

stage. Also it must have a high specificity to minimize false positives that necessitate cost or invasive examination and additional scares the patient and the families needlessly [Ahlquist, 1997]. That one biomarker will accomplish all these criteria will be almost impossible, but the combination of specific markers could have the possibility to meet the condition for a useful screening test in CRC.

This study shows that the ELISA, that detects serum based CCSA-2, is both sensitive and specific for colon cancer. In addition, this is the first time, that CCSA-2 has been detected in the serum from patients with advanced adenomas, confirming tissue data we could found in previous studies in colon polyps [Brunagel et al., 2004]. The serum based ELISA with CCSA-2 antibody demonstrated a sensitivity of 88.8% and considering the entire study population, a specificity of 84.2%.

Three of the colon cancer patients, two had a tumor stage UICC II and one patient UICC III were under the cut off point and therefore considered to be negative for CCSA-2. So far, we have no explanation, why these patients do not appear to express CCSA-2 in the serum. Previously studies have shown that CCSA-2 is expressed in 80% of colon cancer tissues (13). With the presumption that not all colon cancers may be express the NMP CCSA-2, we understood the limitations and the development of

TABLE III. Area Under the ROC Curve

1. ROC analyses for CCSA-2 in separating control individuals (normal colon) from colon cancer patients	
Area	0.9394
Std. error	0.02751
95% Confidence interval	0.8854–0.9933
P-value	<0.0001
Data	
Control	40
Colon cancer patient	27
2. ROC analyses for CCSA-2 in separating control individuals (normal colon) and all other patients from colon cancer patients	
Area	0.8938
Std. error	0.02831
95% Confidence interval	0.8384–0.9493
P-value	<0.0001
Data	
Control and all other patients	127
Colon cancer patient	27

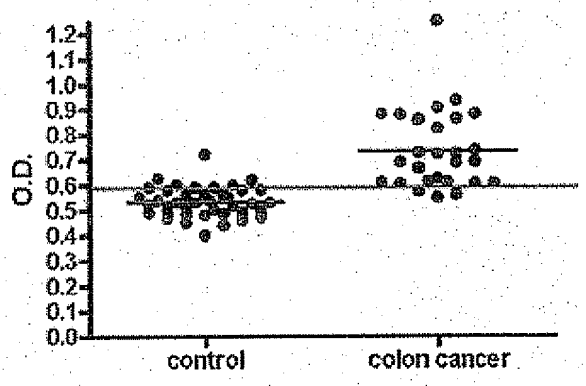


Fig. 2. Serum analysis of CCSA-2 in colon cancer and control. Using the ROC curve a cut off represented by red line of 0.6 OD, results in the optimal balance between sensitivity and specificity.

additional serum marker based on the other identified NMPs CCSA-3,4,5 could close this gap [Leman et al., 2007].

Based on evidence from epidemiological and pathological studies, most sporadic colon cancers are thought to develop from benign adenomas. Presently, there is no clear way of identifying which adenomas will become malignant. There is consent, that progression is associated with severe dysplasia, patient age, size of adenoma, and histological types [O'Brien et al., 1990]. Adenomas that are >1 cm, show severe dysplasia and/or villous architecture are described as advanced.

In previous studies, we demonstrated the expression of CCSA-2 in advanced polyps [Brunagel et al., 2004]. Four serum levels from patients with colon polyps are above the cut off point, three of which have advanced adenomas.

Three normal individuals showed an increased level of CCSA-2 in there serum. Regarding the colonoscopy report, the examination was not difficult and the colon clean. Reviewing the literature 4% of polyps or carcinoma are overseen in a colonoscopy especially in the right colon [Bressler et al., 2004]. In these cases and additionally in the cases with diverticulosis and IBD we can just speculate, if there was something overseen. However, especially in cases where colonoscopy is difficult, a serum marker, which could detect early colon cancer and furthermore advanced adenomas, would be very helpful.

Regarding the other cancers types, 9 patients out of 37 have an expression of CCSA-2 in the serum above the cut off point. Three patients with cholangiocarcinoma (3/6), one patient with lung cancer (1/4), four patients with gastric cancer (4/13), and one patient with hepatocellular carcinoma (1/3). None of the 11 patients with pancreatic cancer had an expression of CCSA-2 above the cut off point. There is no correlation of the tumor stage and the CCSA-2 expression (correlation coefficient (r) = −0.1687, r squared = 0.02847).

So far we have no explanation for the expression of CC2 in other cancer types.

To evaluate the effect of the removal of the colon cancer by surgery on the serum CCSA-2 value, samples were obtained from nine patients after colon cancer surgery 2–9 years after curative surgery. All nine individuals considered to be normal after curative colon cancer surgery.

Additionally, patients with benign inflammatory disease like pancreatitis and gastritis and diverticulitis and patients with IBD were studied. One patient with IBD (1/11), four patients with diverticulosis (4/21), and two patients with benign inflammatory disease (2/9) had CCSA-2 values above the cut off point. We could not observe a correlation between the elevated CCSA-2 levels and the grade of the inflammation.

Further studies are needed to examine the expression on CCSA-2 in other disease. Nevertheless the overall specificity of CCSA-2 is 84.2%, shown it is a specific marker for colon cancer. This is the first study demonstrating the ability of CCSA-2 antibodies to specifically identify colon cancer patients in a clinically applicable test. However, clinical trials need to be performed, for evaluation of the sensitivity

TABLE IV. Specificity/Sensitivity of Blood CCSA-2 Assay

	No. of samples <0.6 OD/total no. samples	Specificity %
Donors	>37/40	>92.5
All populations	>106/127	>84.2
	No. of samples >0.6 OD/total no. samples	Sensitivity %
Colon cancer	>24/27	>88.8

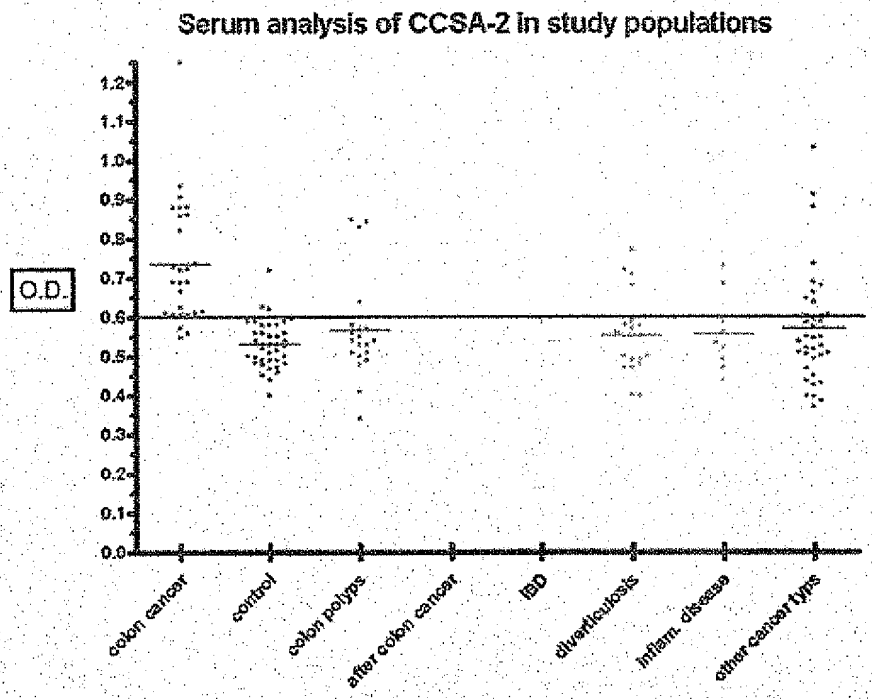


Fig. 3. Serum analysis of CCSA-2 in study populations. Total of 174 serum samples screened for CCSA-2 in indirect ELISA. Cut off value of 0.6 OD. Represented by red line across graph. Line in between the patients groups represents the median value of the group.

and specificity in independent validation studies in a larger population of patients.

ACKNOWLEDGMENTS

This work was supported by a grant from the Else Kröner-Fresenius-Stiftung, the BONFOR Programme of the Medical Faculty, University of Bonn and the Lise-Meitner Fellowship through the Ministerial of Nordrhein-Westfalen, Germany to G. Walgenbach-Brunagel.

REFERENCES

Ahlquist DA. 1997. Fecal occult blood testing for colorectal cancer, Can we afford to do this? Gastroenterol Clin North Am 26:41–55.

Bressler B, Paszat LF, Vinden C, Li C, He J, Rabeneck L. 2004. Colonoscopic miss rates for right-sided colon cancer: A population-based analysis. Gastroenterology 127:452–456.

Brunagel G, Schoen RE, Bauer AJ, Vietmeier BN, Getzenberg RH. 2002a. Nuclear matrix protein alterations associated with colon cancer metastasis to the liver. Clin Cancer Res 8:3039–3045.

Brunagel G, Vietmeier BN, Bauer AJ, Schoen RE, Getzenberg RH. 2002b. Identification of nuclear matrix protein alterations associated with human colon cancer. Cancer Res 62:2437–2442.

Brunagel G, Schoen RE, Getzenberg RH. 2004. Colon cancer specific nuclear matrix protein alterations in human colonic adenomatous polyps. J Cell Biochem 91: 365–374.

Dhir R, Vietmeier B, Arlotti J, Acquafondata M, Landsittel D, Masterson R, Getzenberg RH. 2004a. Early identification of individuals with prostate cancer in negative biopsies. J Urol 171:1419–1423.

Dhir R, Vietmeier B, Arlotti J, Acquafondata M, Landsittel D, Masterson R, Getzenberg RH. 2004b. Early identification of individuals with prostate cancer in negative biopsies. J Urol 171:1419–1423.

Getzenberg RH. 1994. Nuclear matrix and the regulation of gene expression: Tissue specificity. [Review] [48 refs]. J Cell Biochem 55:22–31.

Getzenberg RH, Pienta KJ, Huang EY, Coffey DS. 1991. Identification of nuclear matrix proteins in the cancer and normal rat prostate. Cancer Res 51:6514–6520.

Gevaert K. 1995. New strategies in high sensitivity characterization of proteins separated from 1-D or 2-D gels. In: Atassi MZ, Appella E, editors. Methods in protein structure analysis. New York: Plenum Press. p 15.

Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. 2007. Cancer Statistics, 2007. CA Cancer J Clin 57:43–66.

Konety BR, Getzenberg RH. 1999. Nuclear structural proteins as biomarkers of cancer. J Cell Biochem 32/33:183–191.

Konety BR, Nangia AK, Nguyen TS, Veitmeier BN, Dhir R, Acierno JS, Becich MJ, Hrebinko RL, Getzenberg RH. 1998. Identification of nuclear matrix protein alterations associated with renal cell carcinoma. J Urol 159:1359–1363.

Konety BR, Nguyen TS, Dhir R, Day RS, Becich MJ, Stadler WM, Getzenberg RH. 2000. Detection of bladder cancer using a novel nuclear matrix protein, BLCA-4 [In Process Citation]. Clin Cancer Res 6:2618–2625.

Leman ES, Cannon GW, Trock BJ, Sokoll LJ, Chan DW, Mangold L, Partin AW, Getzenberg RH. 2007. EPCA-2: A highly specific serum marker for prostate cancer. Urology 69:714–720.

Mandel JS. 2005. Screening of patients at average risk for colon cancer. Med Clin North Am 89:43–59, vii.

Myers-Irvin JM, Van Le TS, Getzenberg RH. 2005. Mechanistic analysis of the role of BLCA-4 in bladder cancer pathobiology. Cancer Res 65:7145–7150.

O'Brien MJ, Winawer SJ, Zauber AG, Gottlieb LS, Sternberg SS, Diaz B, Dickersin GR, Ewing S, Geller S, Kasimian D. 1990. The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. Gastroenterology 98: 371–379.

Parkin DM. 2001. Global cancer statistics in the year 2000. Lancet Oncol 2:533–543.

Paul B, Dhir R, Landsittel D, Hitchens MR, Getzenberg RH. 2005. Detection of prostate cancer with a blood-based assay for early prostate cancer antigen. Cancer Res 65: 4097–4100.

Pickhardt PJ, Choi JR, Hwang I, Butler JA, Puckett ML, Hildebrandt HA, Wong RK, Nugent PA, Mysliwiec PA, Schindler WR. 2003. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. N Engl J Med 349:2191–2200.

Pienta KJ, Partin AW, Coffey DS. 1989. Cancer as a disease of DNA organization and dynamic cell structure. Cancer Res 49:2525–2532.

Van Le TS, Myers J, Konety BR, Barder T, Getzenberg RH. 2004. Functional characterization of the bladder cancer marker, BLCA-4. Clin Cancer Res 10:1384–1391.

Walsh JM, Terdiman JP. 2003. Colorectal cancer screening: Scientific review. JAMA 289:1288–1296.

Winawer S, Fletcher R, Rex D, Bond J, Burt R, Ferrucci J, Ganiats T, Levin T, Woolf S, Johnson D, Kirk L, Litin S, Simmang C. 2003. Colorectal cancer screening and surveillance: Clinical guidelines and rationale-Update based on new evidence. Gastroenterology 124:544–560.

EXHIBIT 3

KIAA1199

Analysis of three sequences which correspond to regions of KIAA1199 confirm that KIAA1199 is differentially expressed (higher) in adenoma tissue compared to normal tissues. Expression patterns for the three KIAA1199 sequences are compared in Table 1 and shown in Figures 1, 2 and 3. The results have been obtained by analysing 19 adenoma patients and 30 non-diseased controls.

Table 1: Comparison of expression patterns for KIAA1199 between adenoma and tissues and non-neoplastic tissues.

SeqID	*D*-val	Fold-Δ	Sens/Spec
7	2.39	4.39	88.4
103	1.91	2.17	83
316	2.48	3.88	89.3

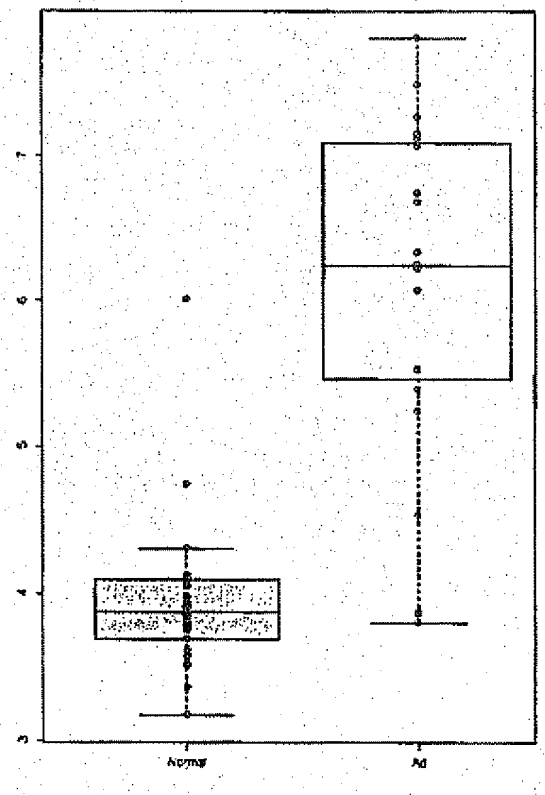


Figure 1. Sequence ID 7 (KIAA1199)

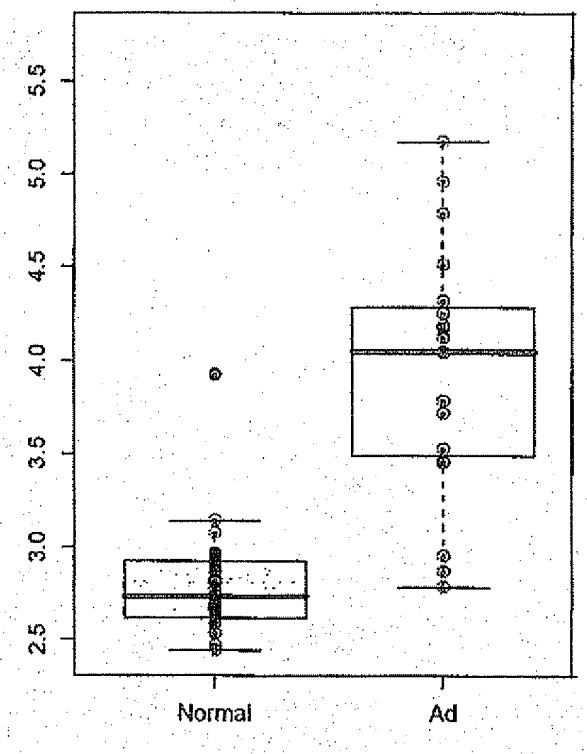


Figure 2: Sequence ID 103 (KIAA1199)

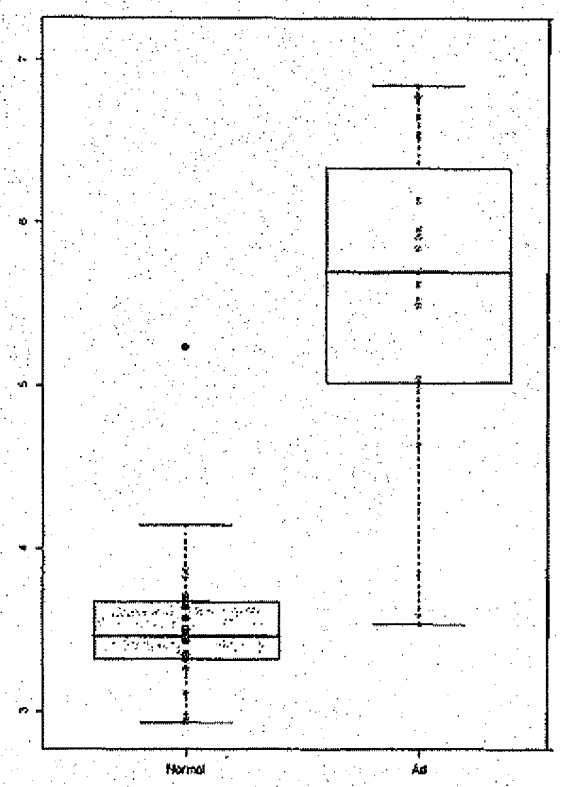


Figure 3: Sequence ID 316 (KIAA1199)

EXHIBIT 4

Transcriptome Profile of Human Colorectal Adenomas

Jacob Sabates-Bellver,[1] Laurens G. Van der Flier,[4] Mariagrazia de Palo,[5] Elisa Cattaneo,[1] Caroline Maake,[2] Hubert Rehrauer,[3] Endre Laczko,[3] Michal A. Kurowski,[9] Janusz M. Bujnicki,[9] Mirco Menigatti,[7] Judith Luz,[8] Teresa V. Ranalli,[6] Vito Gomes,[6] Alfredo Pastorelli,[5] Roberto Faggiani,[5] Marcello Anti,[5] Josef Jiricny,[1] Hans Clevers,[4] and Giancarlo Marra[1]

[1]Institute of Molecular Cancer Research, [2]Institute of Anatomy, and [3]Functional Genomic Center, University of Zurich, Zurich, Switzerland; [4]Hubrecht Institute, Netherlands Institute for Developmental Biology, Utrecht, the Netherlands; [5]Gastroenterology Unit and [6]Pathology Department, Belcolle Hospital, Viterbo, Italy; [7]Institute of Biochemistry and Genetics and [8]Division of Medical Genetics, University of Basel, Basel, Switzerland; and [9]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Warsaw, Poland

Abstract

Colorectal cancers are believed to arise predominantly from adenomas. Although these precancerous lesions have been subjected to extensive clinical, pathologic, and molecular analyses, little is currently known about the global gene expression changes accompanying their formation. To characterize the molecular processes underlying the transformation of normal colonic epithelium, we compared the transcriptomes of 32 prospectively collected adenomas with those of normal mucosa from the same individuals. Important differences emerged not only between the expression profiles of normal and adenomatous tissues but also between those of small and large adenomas. A key feature of the transformation process was the remodeling of the Wnt pathway reflected in patent overexpression and underexpression of 78 known components of this signaling cascade. The expression of 19 Wnt targets was closely correlated with clear up-regulation of KIAA1199, whose function is currently unknown. In normal mucosa, KIAA1199 expression was confined to cells in the lower portion of intestinal crypts, where Wnt signaling is physiologically active, but it was markedly increased in all adenomas, where it was expressed in most of the epithelial cells, and in colon cancer cell lines, it was markedly reduced by inactivation of the β-catenin/T-cell factor(s) transcription complex, the pivotal mediator of Wnt signaling. Our transcriptomic profiles of normal colonic mucosa and colorectal adenomas shed new light on the early stages of colorectal tumorigenesis and identified KIAA1199 as a novel target of the Wnt signaling pathway and a putative marker of colorectal adenomatous transformation.
(Mol Cancer Res 2007;5(12):1263–75)

Introduction

In developed countries, sporadic adenomatous colorectal polyps are found in roughly one third of asymptomatic adults below the age of 50 who undergo colonoscopy. Depending on their characteristics (multiplicity, size, histologic features, and degree of dysplasia), these lesions can be associated with a substantial risk of recurrence (up to 60% at 3 years) and the development of advanced neoplastic disease (reviewed in ref. 1 and references therein). It has been estimated that 15% of all adenomas measuring ≥1 cm will progress to carcinomas within 10 years of their detection (2).

Although adenomatous polyps are not the only precancerous lesions in the colorectum, they are the most common, and they are the precursors of most of the cancers in this organ. In these neoplasms, the transformation process begins in the epithelial crypts and seems to result from qualitative, quantitative, and spatial subversion of the Wnt signaling pathway, the physiologic regulator of epithelial homeostasis (3-5). This adenoma-carcinoma pathway of tumorigenesis is characterized by mutations involving various components of this pathway (e.g., APC, whose germ-line mutations are responsible for familial adenomatous polyposis; CTNNB1, which encodes a subunit of the cadherin protein complex known as β-catenin; and Axin, the gene encoding a multidomain scaffold protein that is essential for β-catenin degradation). The result of these mutations is an accumulation of β-catenin, first in the cytoplasm and then in the nucleus, where it associates with DNA-binding proteins of the T-cell factor (TCF)/lymphoid enhancer factor family, transforming them from transcriptional repressors into transcriptional activators that affect the expression of numerous genes involved in epithelial homeostasis.

Although the key role played by adenomatous polyps in colorectal tumorigenesis is widely acknowledged, the gene expression changes that trigger or accompany their development

Received 6/9/07; revised 7/28/07; accepted 8/2/07.
Grant support: Zurich Cancer League and Swiss National Science Foundation.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Note: Supplementary data for this article are available at Molecular Cancer Research Online (http://mcr.aacrjournals.org/).
J. Sabates-Bellver and L.G. Van der Flier contributed equally to this work.
Requests for reprints: Giancarlo Marra, Institute of Molecular Cancer Research, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland. Phone: 41-044-635-3472; Fax: 41-044-635-3484. E-mail: marra@imcr.uzh.ch

doi:10.1158/1541-7786.MCR-07-0267

have never been comprehensively studied. We therefore conducted a transcriptomic analysis of prospectively collected colorectal adenomas using a standardized oligonucleotide microarray covering the entire human genome. This study not only provided new information that is fundamental for future molecular characterization of these precancerous lesions but also allowed us to identify a putative marker of colorectal tumorigenesis.

Results

The focus of our study was the adenoma-adenocarcinoma pathway of colorectal carcinogenesis, which is closely linked to deregulation of the Wnt signaling pathway. To gain insight into the early steps of this process, we confined our investigation exclusively to sporadic, pedunculated colorectal adenomas (type 0-Ip of the Paris classification; ref. 6). Nonpolypoid and sessile polypoid lesions were not included because in some cases their transformation is believed to proceed along nonadenomatous pathways (7). Details on our case selection criteria are provided in Materials and Methods.

Thirty-two pedunculated adenomatous polyps, each with matched samples of normal mucosa, were prospectively collected from 28 patients (Table 1). The total number of synchronous and previously excised adenomas was <3 in 18 of 28 patients and 3 to 15 in the remaining 10. In this latter subgroup, the absence of APC- or MYH-associated multiple adenomatosis had been confirmed by genetic testing of lymphocyte DNA. Histologic analysis of one polyp (case NM) revealed superficial infiltration of the submucosa, but this case was not excluded because the region sampled for microarray analysis was clearly adenomatous. (As noted below, this finding was consistent with the results of hierarchical cluster analysis shown in Supplementary Fig. S1.).

Analysis of microarray data for the 32 adenoma/normal mucosa tissue pairs revealed that 31,033 of the probes were expressed in one or both of the tissue groups. The normal tissues were effectively segregated from the adenomas in four unsupervised analyses of the expression levels of these genes [hierarchical clustering, principal component analysis (PCA), correlation analysis, and correspondence analysis (CA); see Materials and Methods for details; Fig. 1]. In a separate

Table 1. Characteristics of the 28 Patients with Adenomatous Polyps Included in the Study

Patient	Age (y)	Sex	Colon segment involved	Maximum adenoma diameter (mm)	Microscopic appearance	Highest degree of dysplasia in the adenoma*	Degree of dysplasia at sampling site*	No. adenomas at study colonoscopy[c]	No. previously excised adenomas[b]	No. previous and/or synchronous hyperplastic polyps	Familiarity for colorectal cancer (relative, onset age)
GL[x]	49	M	D/S	10/10	T-V/T-V	H/L	H/L	9	4[k,{]	15[**]	Mother, 70
PR	74	F	S	20	T-V	H	H	2	—[cc]	0	No
PC[x]	69	M	S/S	10/20	T/T-V	H/H	H/H	10	—	1	No
FP[x]	57	M	S/S	15/30	T/T-V	H/H	H/L	1	—	1	Mother, 69
CD[x]	71	M	T/R	15/10	T/T	H/L	H/L	2	7[k]	2	No
MA	65	M	R	15	T-V	L	L	2	—	0	No
ME	63	M	R	15	T-V	L	L	9	—	1	Father, 79
RA	64	F	A	15	T	L	L	1	7[k]	0	Sister, 68
PR	72	M	R	40	T-V	H	H	5	—	0	No
SD	56	M	A	15	T	H	H	1	1[k]	0	Mother, 83; sister, 87
MP	38	M	S	15	T-V	L	L	2	—	0	Father, 79
MP	61	M	S	20	T	L	L	3	—	2	no
LG	41	M	R	20	T-V	H	L	5 (2 serrated)	—	0	Father, 60
LS	45	M	S	20	T	H	H	1	—	1	Father, 60
BG	58	M	D	15	T-V	L	L	2	—	1	No
PL	69	F	S	15	T-V	L	L	2	—	0	No
SMA	52	F	S	30	T	H	H	2	—	1	No
MR	58	F	D	20	T	L	L	2	—	0	No
GN	69	M	R	40	T	H	H	2	—	0	No
BA	59	M	S	30	T	L	L	6	—	0	No
PF	56	M	S	30	T	L	L	2	—	0	No
RC	55	F	A	30	T-V	L	L	12	3[k,{]	1	No
TMA	58	F	S	10	T	L	L	1	—	0	Mother, 85
NM	52	M	R	35	T-V	T1[bb]	H	1	—	0	No
MA	83	M	S	10	T-V	H	H	2	—	1	No
MM	50	M	S	30	T-V	H	H	2	—	0	Father's brother, 65
NF	79	M	A	20	T-V	L	L	2	—	0	Mother, 70
PN	67	F	S	15	T-V	L	L	1	—	1	No

Abbreviations: M, male; F, female; A, ascending colon; T, transversum; D, descending colon; S, sigmoid colon; R, rectum; T, tubular; T-V, tubulovillous; L, low-grade dysplasia; H, high-grade dysplasia.

*Low-grade versus high-grade dysplasia as defined by the WHO classification of tumors of the digestive system, editorial and consensus conference in Lyon, France, November 6-9, 1999, IARC.

cThis number includes the adenoma(s) subjected to microarray analysis.

bTotal number of adenomas detected and excised during previous colonoscopies.

xTwo adenomas from these patients were analyzed.

kThese cases were considered as recurrent adenomas for the CCA.

{The index colonoscopy was done in a different center about 10 y before the study colonoscopy.

**Hyperplastic polyposis.

ccNo previous colonoscopies.

bbSuperficial submucosal invasion (T1). The tissue collected for microarray came from the adenomatous portion of the polyp.

analysis, these two tissue groups were also unequivocally distinguished from a previously described set of 25 colon cancers (8), which we reanalyzed for this study with the same microarray used to characterize the adenomas and normal mucosa (Supplementary Fig. S1).

Almost half of the expressed probes (15,059 of 31,033) displayed significant expression changes in adenomas. Those with fold changes ≥2 (1,190 probes up-regulated and 2,469 down-regulated in adenomas) were subjected to gene ontology analysis to identify the biological processes involved in the transition of normal mucosa to adenoma. The most significant results of this analysis are listed in Supplementary Table S1. The processes that were most markedly overrepresented among genes that were up-regulated in adenomas included mitosis, DNA replication, and spindle organization. Down-regulated genes were predominantly involved in host immune defense, inorganic anion transport, organ development, and inflammatory response, although a small group of genes involved in the latter process was up-regulated in adenomas (Supplementary Fig. S2).

We then analyzed the transcript levels of 319 genes believed to be components of the complex Wnt signaling pathway (Supplementary Table S2). Sixty-six of these genes (21%) were not expressed in either the normal or adenomatous tissue, and 34% were expressed similarly in both tissue groups. The remaining 144 genes displayed significantly altered expression in adenomas, and 78 of 144 displayed fold changes of ≥2.

A supervised extension of CA (9), canonical CA (CCA), was then used to identify possible correlations between gene expression patterns and clinical or pathologic variables. Four of the variables considered (adenoma diameter, colon segment of origin, degree of dysplasia, and adenoma recurrence; see Table 1) were clearly associated with distinct clusters of expression profiles (Fig. 2, variables in A and clusters for adenoma diameter in B; more details in the legend to this figure). The profile of adenomas measuring >20 mm could be easily distinguished from those of smaller (≤20 mm) adenomas. As shown by CCA and visualized on the corresponding CCA score plot (Fig. 2B), the centers of the three adenoma size clusters are distributed along the principal CCA axis (the vertical axis in Fig. 2B, the most important axis of separation of the expression profiles) in a definite order, with increasing diameters corresponding to progressively higher CCA scores. The variable large adenoma diameter was closely correlated with the vertical CCA axis (i.e., its vector "d>20mm" in Fig. 2A is almost parallel to this axis). It is interesting to note that the same correlation can be observed for the variable high-degree dysplasia (i.e., represented in Fig. 2A by vector "Hd"). This finding confirms the expected correlation between larger diameters and higher-degree dysplasia.

The CCA plot of the 11,709 modeled probes (loading plot, not shown) suggested that the distinction between the three size groups of adenomas is due to a complex network of relatively small changes in the expression of numerous genes (as opposed to marked changes involving a limited number of genes). Nevertheless, to maximize the use of the extensive data sets, we selected the 500 probes with the highest loading scores along the CCA axis 1 and isolated a set of genes whose expression changes displayed significant positive or negative correlation with adenoma size (Supplementary Table S3). Although their association with adenomas must be validated in a larger series, these are the expression changes most likely to play causal roles in the progression of these tumors.

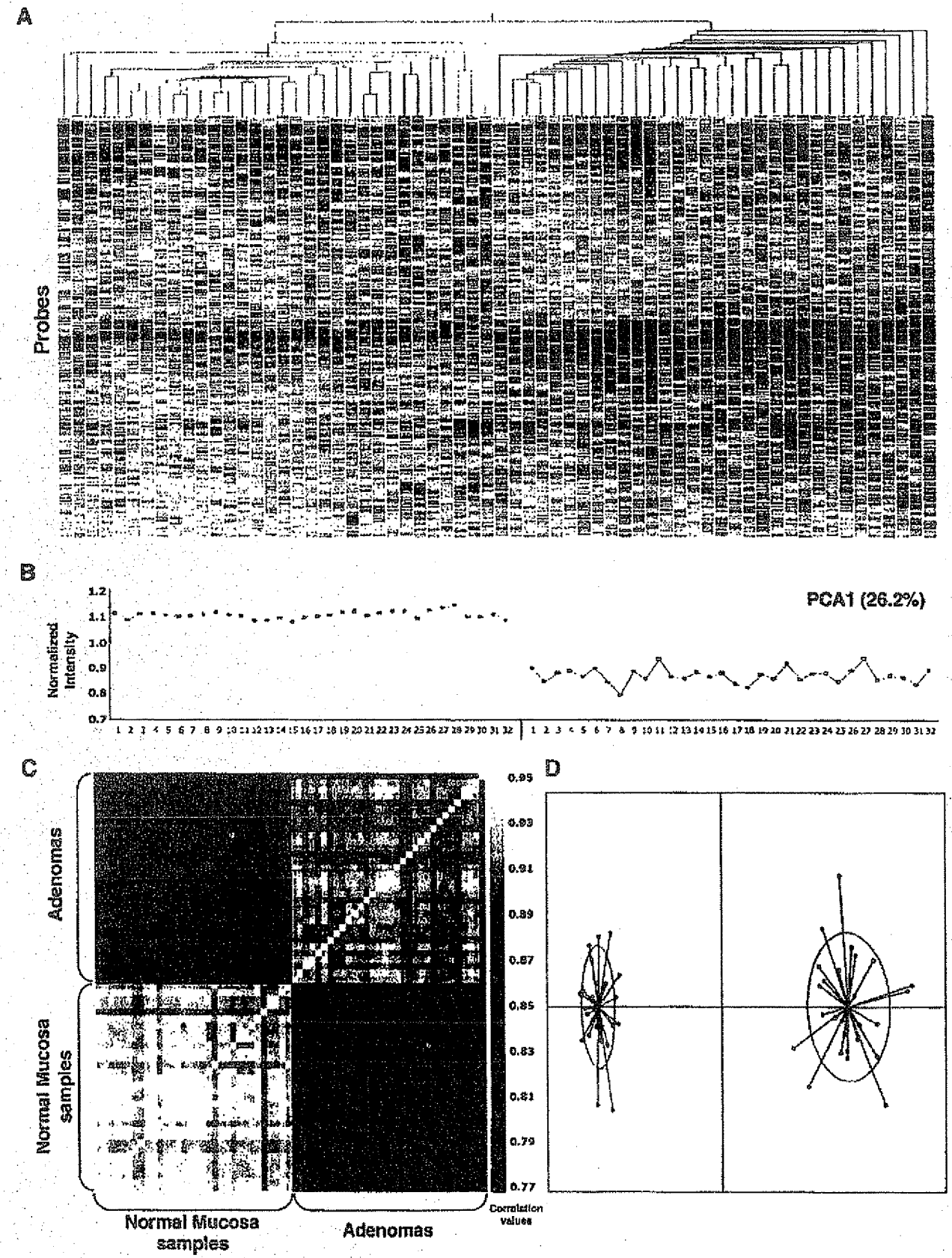
It should be mentioned that normal mucosa from the sigmoid colon had an expression profile that differed significantly from that of tissues from other colon segments (Fig. 2A). This finding will be explored in a future study conducted on a large series of normal mucosa samples from different colorectal segments.

The transcriptional profile of the 32 adenomas was thoroughly analyzed to identify genes likely to be involved in the development and evolution of these lesions. One of the first features that attracted our attention was the marked up-regulation of KIAA1199 (Supplementary Table S4), a gene encoding a protein with unknown function. Its overexpression was striking in all colorectal adenomas we examined (average increases of 54.8-fold compared with normal mucosa) and in carcinomas (8). These findings were fully confirmed by real-time reverse transcription-PCR analysis of RNA extracted from samples used for the microarray study and from additional samples collected after the present study was completed (Supplementary Fig. S3).

In light of these findings, it was natural to wonder whether KIAA1199 might be a novel positively regulated target of Wnt signaling, which is characteristically deregulated in colorectal tumors. Previous microarray studies indicated that genes coregulated at the transcriptional level under different conditions tend to be involved in the same processes and pathways, and the analysis of transcriptional coexpression has been used to predict the function of novel genes (10-12). Therefore, we conducted a search for known Wnt targets (listed in Supplementary Table S5) among the genes whose expression patterns in all the tissue samples significantly correlated with those of KIAA1199. (The procedure used in this analysis is summarized in Materials and Methods and Supplementary Fig. S4.) Forty-nine percent of the known Wnt targets that were overexpressed in our adenoma samples had expression patterns that were positively correlated with that of KIAA1199 (Fig. 3A and B) as opposed to only 7.9% of the overexpressed genes that are not considered Wnt targets ($P < 0.0001$).

Evidence of the potential involvement of KIAA1199 in the Wnt signaling pathway had also emerged from another study by our group (13). A combined analysis of microarray data of tissues and cell lines placed KIAA1199 at the top of a list of genes [Supplementary Table S1 of ref. 13] that were up-regulated in colorectal adenomas and down-regulated in colon cancer cell lines that had undergone stable transfection with doxycycline-inducible forms of dominant-negative TCF1 or TCF4 to suppress Wnt signaling (14, 15). In the present study, KIAA1199 was also found to be markedly down-regulated in LS174T colon cancer cells in which Wnt signaling had been blocked by the induction of h-catenin small interfering RNA or NH_2-terminal–deleted TCF4 (15, 16). The dramatic decrease in KIAA1199 mRNA levels associated with this inhibition of the Wnt pathway was confirmed by Northern blotting (Fig. 3C).

In general, Wnt target genes are expressed predominantly in the proliferating compartment of normal intestinal crypts (lower portion), and their expression is appreciably increased in


A
Probes
B
Normalized Intensity
1.2
1.1
1.0
0.9
0.8
0.7
PCA1 (26.2%)
C
Adenomas
Normal Mucosa samples
Normal Mucosa samples
Adenomas
0.95
0.93
0.91
0.89
0.87
0.85
0.83
0.81
0.79
0.77
Correlation values
D

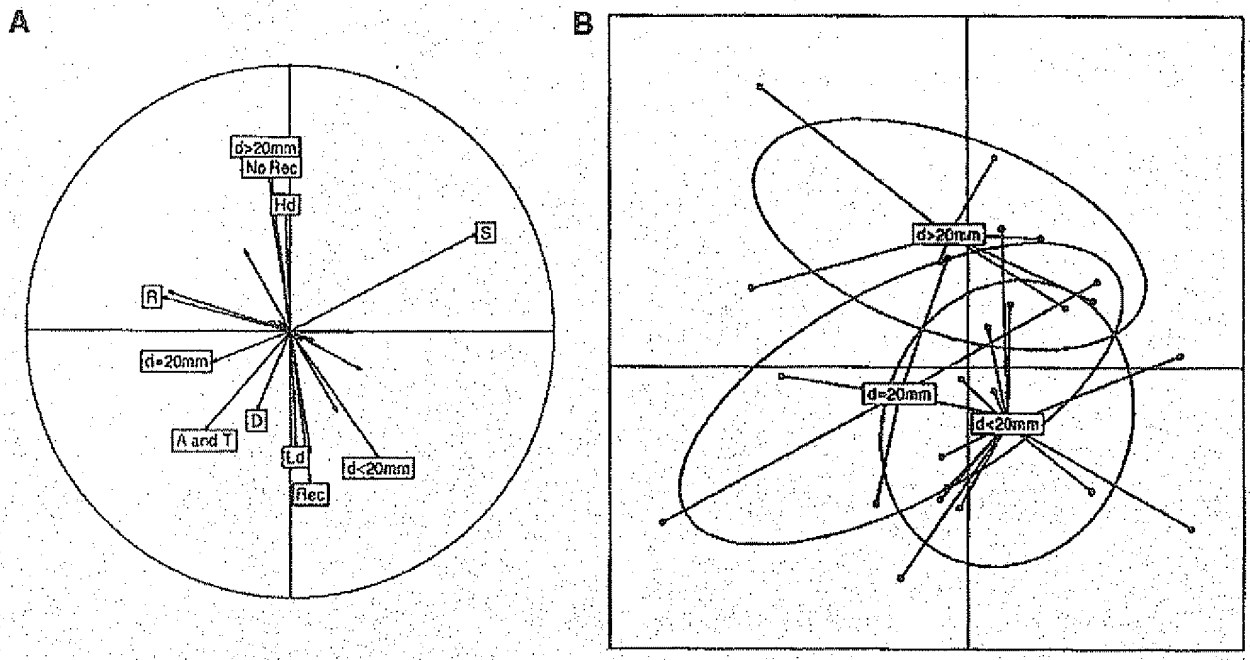


FIGURE 2. Clinical/pathologic variables that correlate with distinct gene expression profiles. The panels summarize the most important results of the CCA of mRNA intensity log-ratio values (adenoma: normal) of expressed genes. For clarity, CCA axis 1 has been drawn vertically in both panels. A. Correlation between specific clinical/pathologic variables (adenoma diameter, colon segment of origin, degree of dysplasia, and adenoma recurrence) and clusters of differential gene expression profiles (coded as log-ratio profiles), such as those shown in B. Each vector represents a specific value for a given variable (e.g., adenoma diameter of >20 mm and high-degree dysplasia) and points toward the center of the profile cluster correlated with the clinical/pathologic characteristic it represents. If the centers for each specific value are separated, the corresponding vectors point in distinct directions; otherwise, they are directed toward the same point. In the former case, the represented variable can be assumed to be significantly correlated with the profiles; in the latter case, there is no correlation. The length of the vector reflects the strength of the correlation: those approaching the circumference of the correlation circle, which represents a correlation value of 1, indicate stronger correlation than shorter vectors (correlation closer to 0). d, diameter; Hd, high-degree dysplasia; Ld, low-degree dysplasia; A, ascending colon; T, transverse colon; D, descending colon; S, sigmoid colon; R, rectum; Rec, recurrent adenomas; no Rec, no recurrent adenomas. Unlabeled vectors are related to variables that were not clearly associated with any distinct cluster of expression profiles. Larger adenomas were predictably associated with high-degree dysplasia. In contrast, their association with nonrecurrence was unexpected and probably due to the fact that patients who had already undergone endoscopic polypectomy (i.e., those with recurrence) presented relatively recent-onset (consequently, smaller) polyps at the study colonoscopy. B. CCA score plot with samples grouped by adenoma diameter. Each of the three size-related groups is delimited by an ellipse with the center labeled. The ellipse representing the adenomas measuring >20 mm in diameter shows very little overlap with those of the other two groups (adenomas with diameters of 20 mm and those with diameters of <20 mm).

adenomatous glands (15). Our analysis of human tissues with preserved architecture indicated that these are also attributes of KIAA1199. In in situ hybridization studies, KIAA1199 mRNA was detectable only in the lower portion of normal colonic epithelial crypts (Fig. 4A and B), and its expression levels were much higher in dysplastic glands (Fig. 4C). These patterns were confirmed at the protein level by immunohistochemistry done with an antibody raised in our laboratory (Fig. 4D-J). This analysis also revealed that the KIAA1199 is a cytoplasmic protein whose expression is most intense near the cell membrane, particularly on the luminal side of the dysplastic cell multilayer (Fig. 4F-J).

FIGURE 1. Unsupervised analyses of microarray data. A. Hierarchical clustering analysis. The 64 tissue samples represented on the X axis include 32 normal mucosal samples (green branches) and 32 adenomas (red branches). Each probe plotted on the Y axis is color coded to indicate the level of expression of the gene relative to its median expression level across the entire tissue sample set (blue, low; red, high). In the adenoma dendrogram, branches representing individual samples and small groups merge at higher levels than those of the normal mucosa dendrogram, reflecting lower-level correlation (i.e., higher variability among the adenoma specimens). B. PCA. Profile plot of the normalized first principal component (PCA1) across the 64 specimens (green dots, normal mucosa; red dots, adenomas). The two tissue groups differ significantly in terms of PCA1 ($P < 0.0001$), which accounted for 26% of the total variance. Note the higher variability of the PCA1 values in the adenoma group (higher fluctuation). C. Correlation analysis. Tile plot visualization of the pairwise correlations of the samples. Correlation values are indicated on the grayscale column (white > black: high > low). High correlation is observed among the samples within each group (top right quadrant, adenomas; bottom left quadrant, normal mucosa), although the adenomas displayed somewhat greater diversity (i.e., on the whole, the gray tones in the top right quadrant are darker than those in the bottom left quadrant). Top left and bottom right quadrants, normal and adenoma samples are poorly correlated. However, samples from the same patient generally showed higher correlation than that observed between normal and adenoma samples from different patients (bright pixels on the secondary diagonals in the top left and bottom right quadrants). This finding probably reflects the strong influence of several factors, including the individual genetic background and lifestyle and the fact that the normal and adenomatous tissues from a given patient were from the same colon segment. D. CA of mRNA log(intensity) values of expressed genes from 27 of the 32 tissue pairs (green dots, normal mucosa; red dots, adenoma). The other five pairs were excluded from this analysis because one of the two samples behaved as an outlier. Limiting our analysis to the more homogeneous pairs facilitated the comparison of the gene expression profiles for the two tissue groups and allowed more reliable identification of clinical/pathologic variables associated with profile scatter (see Fig. 2). The areas delimited by the ellipses represent 95% of the estimated binormal distribution of the sample scores on the first and second CA axes. The map of the sample scores on the first two axes shows that CA efficiently discriminates between normal and adenoma samples. Higher variability is evident in the adenoma group, where the samples are more widely dispersed.

Discussion

Adenomatous colorectal polyps are one of the most common human tumors and the most frequent precancerous lesions in the colorectum, but their transcriptome has been only partially analyzed, and the data are generally based on a limited number of cases (17-20). We attempted to fill this gap by doing a comprehensive whole-genome microarray analysis of a large, highly homogenous set of adenomas that was collected prospectively.

A comparison of the transcriptomes of adenomatous polyps and segment-matched samples of normal colorectal mucosa revealed evidence of broad-scale remodeling. As a starting point for future verification studies, we have drawn up a list of 478 genes that were significantly up-regulated ($n = 153$) or down-regulated ($n = 325$) in the adenomatous tissues (fold changes of ≥ 4; Supplementary Table S4). Space constraints preclude more than a cursory examination of this list, but we have highlighted in Table 2 certain aspects that we feel are particularly interesting in terms of their relevance to the process of adenoma formation. For instance, transcription regulation seems to be extensively modified. Twenty-nine molecules involved in this process were expressed in adenomas at levels

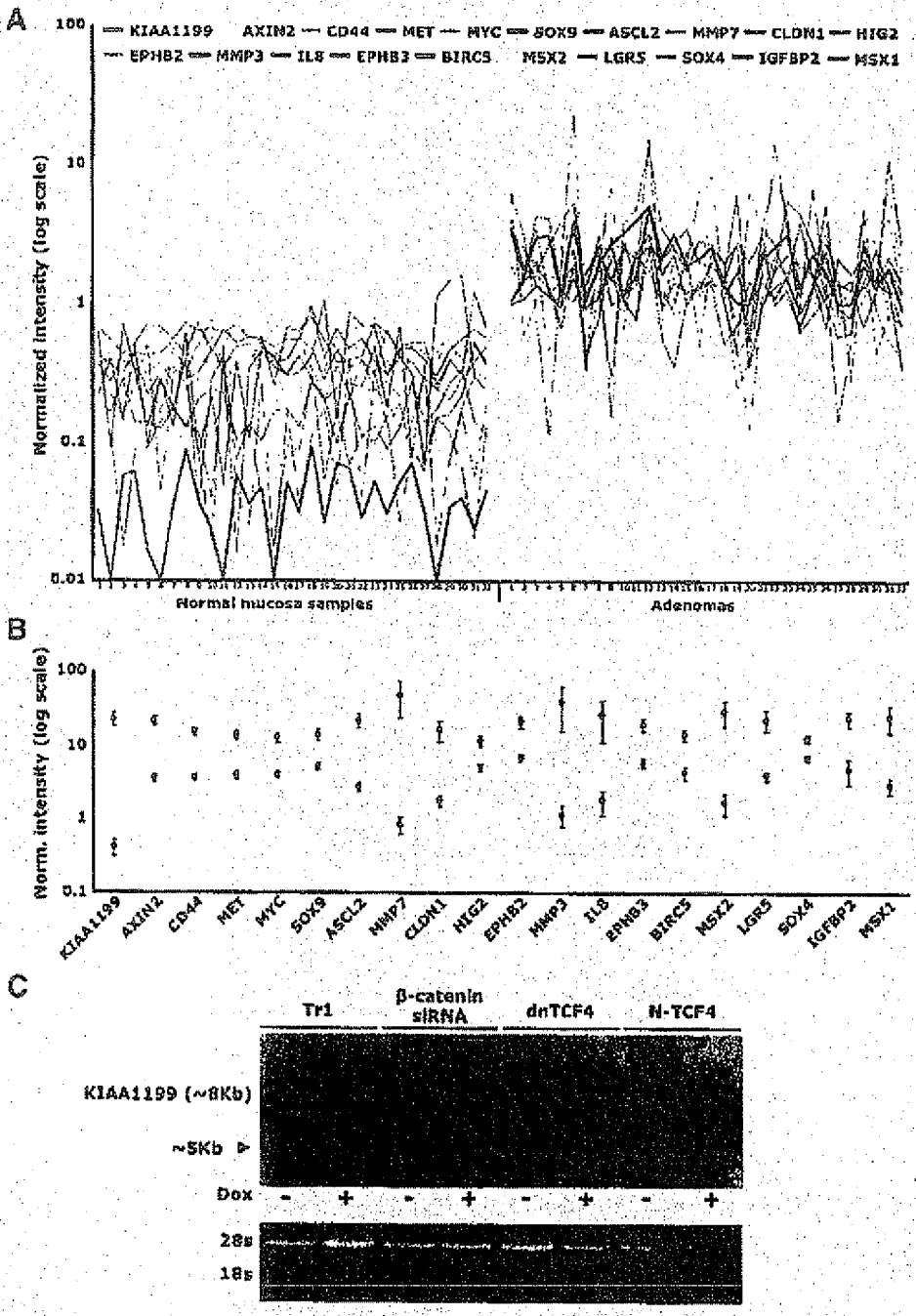


FIGURE 3. KIAA1199 is a putative target of Wnt signaling. A, Degree of correlation between the expression of KIAA1199 mRNA and that of 19 known Wnt signaling target genes identified with the procedure described in Materials and Methods, Results, and Supplementary Fig. S4. For each of the 20 genes, the graph shows the normalized intensity of expression level (plotted on the Y axis) in each of the 32 adenomas and corresponding samples of normal mucosa (X axis). B, Mean expression of each gene in normal mucosa (green dots) and adenomas (red dots). Bars, confidence interval. C, Northern blot showing reduced KIAA1199 expression in LS174T cells following doxycycline-mediated induction of h-catenin small interfering RNA, dominant-negative TCF4 (dnTCF4), or NH_2-terminal–deleted TCF4 (N-TCF4). The ~8-kb band corresponds to full-length KIAA1199 mRNA. The lower band (~5 kb) may represent an alternative form of this mRNA. Dox, cell transfectants grown in the presence or absence of doxycycline; Tr1, a parental clone (i.e., cells expressing the repressor protein modified by doxycycline but not transfected with h-catenin small interfering RNA, dominant-negative TCF4, or NH_2-terminal–deleted TCF4) used as a control of doxycycline exposure. Bottom, ethidium bromide–stained agarose gel as a loading control.

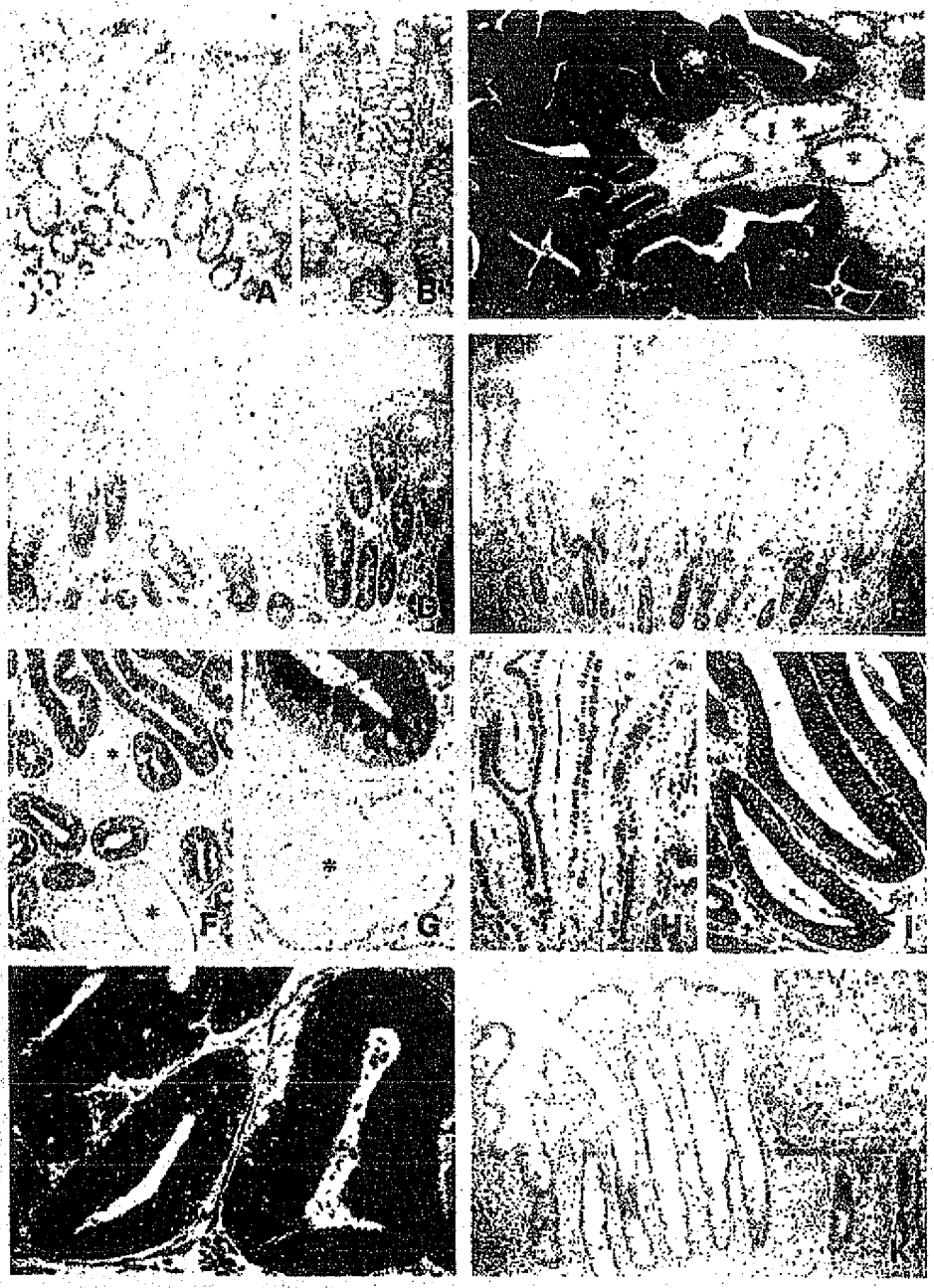


FIGURE 4. Expression of KIAA1199 mRNA and protein in normal intestinal mucosa and colorectal tumors. In situ hybridization studies (A-C) localized KIAA1199 mRNA expression to the lower portion of normal epithelial crypts (A and B) and revealed that expression is markedly up-regulated in colorectal tumors (C). Asterisk, note the different levels of expression in tumor glands and normal crypts. D. KIAA1199 protein expression is also limited to the lower half of the normal colonic crypts, and a similar pattern is observed in the ileal mucosa (E), where the protein is expressed only in the crypts (not in the villi). In F and G, adenomatous crypts with low-grade dysplasia present increased expression of KIAA1199, particularly in the cytoplasm facing the crypt lumen, and in and around the mucin vacuoles of goblet cells (note the striking difference with goblet cells of normal crypts in both panels). The expression pattern changes dramatically during the transition from low-grade dysplasia with goblet cell differentiation (H) to high-grade dysplasia in which this differentiation is no longer apparent. J. In more advanced colon tumors, KIAA1199 overexpression is maintained. Note that, in I and J, the expression of KIAA1199 protein (like that of KIAA1199 mRNA; C) is highest in the luminal portion of the dysplastic glands (arrowheads, multilayer of unstained nuclei occupying more than the basal half of the dysplastic epithelium). K. Normal mucosa, with the corresponding tumor in the inset. Negative control: KIAA1199 antibody preabsorbed with the peptide used to immunize rabbits.

>4 higher or lower than those observed in the normal mucosa, but there were also several smaller changes in this category (Supplementary Table S6) that might also have dramatic effects on gene expression. Several other alterations reported in Table 2 are noteworthy in terms of their potential effect on cell proliferation, differentiation, apoptosis, and cell adhesion: (a) up-regulation of four members of the REG (regenerating) family of genes (21, 22), which would lead to increased tissue mitogen expression; (b) up-regulation of LCN2 (23) and down-regulation of ZFHX1B/SIP-1 (24) in the absence of significant changes in the expression of the epithelial cadherin CDH1 (E-cadherin), which would prevent or delay the epithelial-mesenchymal transition [changes were also noted in the expression of other cell adhesion genes of the cadherin and claudin families, including the striking overexpression of the placental cadherin gene CDH3, which is associated with early events in the transformation process (25, 26)]; (c) down-regulation of ZFHX1B/SIP-1 and Max dimerization protein 1 (MXD1/MAD1; decreased only 3.3-fold and therefore not listed in Table 2; refs. 27, 28) and overexpression of the RTEL1 helicase, which should facilitate telomere elongation (29); (d) alterations that would diminish apoptosis [e.g., overexpression of the decoy receptor for Fas ligand, TNFRSF6B, which is reportedly coregulated with RTEL1 on chromosome 20q13.3

Table 2. Genes Most Likely to be Involved in the Development and Evolution of Colorectal Adenomas (A Subset of Genes Listed in Supplementary Table S4) Subdivided by Gene Ontology Category

Gene symbol	Gene name	Fold differences*	
		E	I
Regulation of transcription			
NLF1	Nuclear localized factor 1	33.1	
FOXQ1	Forkhead box Q1	24.4	
MSX2	Msh homeobox homologue 2	22.2	
ASCL2	Achaete-scute complex-like 2	17.3	
MSX1	Msh homeobox homologue 1	8.5	
IRX3	Iroquois homeobox protein 3	8.4	
GRHL3	Grainyhead-like 3	7.9	
TRIM29	Tripartite motif-containing 29	7.4	
ETV4	Ets variant gene 4 (E1A enhancer binding protein, E1AF)	5.4	
ARNTL2	Aryl hydrocarbon receptor nuclear translocator-like 2	5.3	
TEAD4	TEA domain family member 4	5.2	
SP5	Sp5 transcription factor	5.2	
HES6	Hairy and enhancer of split 6	4.6	
TBX3	T-box 3	4.6	
NFE2L3	Nuclear factor (erythroid-derived 2)-like 3	4.3	
GRHL1	Grainyhead-like 1	4.2	
FEV	FEV (ETS oncogene family)		15.1
SPIB	Spi-B transcription factor		13.2
NEUROD1	Neurogenic differentiation 1		10.6
MEIS1	Meis1, myeloid ecotropic viral integration site 1		7.1
NR3C1	Nuclear receptor subfamily 3, group C, member 1		5.9
NR5A2	Nuclear receptor subfamily 5, group A, member 2		5.6
THRB	Thyroid hormone receptor, b		5.2
ZNF483	Zinc finger protein 483		5.1
ZFHX1B	Zinc finger homeobox 1b (SIP-1)		4.8
MEOX2	Mesenchyme homeobox 2		4.7
HOXD10	Homeobox D10		4.6
MAF	v-maf musculoaponeurotic fibrosarcoma oncogene		4.5
SOX10	SRY (sex determining region Y)-box 10		4.2
Cell proliferation/differentiation/apoptosis			
REG1B	Regenerating islet-derived 1b	75.8	
REG3A	Regenerating islet-derived 3a	29.5	
TACSTD2	Tumor-associated calcium signal transducer 2	21.4	
IL-8	Interleukin-8	14.7	
SERPINB5	Serpin peptidase inhibitor, clade B, member 5 (Maspin)	13.8	
REG1A	Regenerating islet-derived 1a	8.2	
FAIM2	Fas apoptotic inhibitory molecule 2	7.5	
DUSP4	Dual specificity phosphatase 4	7.4	
REG4	Regenerating islet-derived family, member 4	6.8	
PHLDA1	Pleckstrin homology-like domain, family A, member 1	6.0	
LCN2	Lipocalin 2 (oncogene 24p3)	5.7	
RTEL1	Regulator of telomere elongation helicase 1	5.6	
TGFBI	Transforming growth factor, b induced	5.2	
IGFBP2	Insulin-like growth factor binding protein 2	4.8	
TDGF1	Teratocarcinoma-derived growth factor 1	4.7	
TNFRSF6B	Tumor necrosis factor receptor superfamily, member 6b, decoy	4.5	
DMBT1	Deleted in malignant brain tumors 1	4.2	
TNFRSF10C	Tumor necrosis factor receptor superfamily, member 10c, decoy	4.1	
ANGPTL1	Angiopoietin-like 1 (Angioarrestin)		24.9
CDKN2B	Cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)		14.9
GPM6B	Glycoprotein M6B		11.5
ANK2	Ankyrin 2		9.8
UNC5C	Unc-5 homologue C		7.4
HPGD	Hydroxyprostaglandin dehydrogenase 15-(NAD)		6.1
CPNE8	Copine VIII		5.5
FAIM3	Fas apoptotic inhibitory molecule 3		5.4
IL6R	Interleukin-6 receptor		4.8
TUSC3	Tumor suppressor candidate 3		4.7
DUSP1	Dual specificity phosphatase 1		4.7
RERG	RAS-like, estrogen-regulated, growth inhibitor		4.6
NDN	Necdin		4.5
IGF1	Insulin-like growth factor 1 (somatomedin C)		4.0
Cell adhesion			
CDH3	Cadherin 3, type 1, P-cadherin	81.7	
CLDN2	Claudin 2	16.1	
CLDN1	Claudin 1	9.0	
DSG3	Desmoglein 3	7.3	

(Continued on the following page)

Table 2. Genes Most Likely to be Involved in the Development and Evolution of Colorectal Adenomas (A Subset of Genes Listed in in Supplementary Table S4) Subdivided by Gene Ontology Category (Cont'd)

Gene symbol	Gene name	Fold differences*	
		E	I
DSG4	Desmoglein 4	5.9	
CLDN8	Claudin 8		25.8
CDH19	Cadherin 19, type 2		8.3
CEACAM7	Carcinoembryonic antigen-related cell adhesion molecule 7		8.3
CLDN23	Claudin 23		8.0
NRXN1	Neurexin 1		7.1
PCDH19	Protocadherin 19		6.8
NLGN4X	Neuroligin 4, X-linked		6.0
TNXB	Tenascin XB		5.6
MUCDHL	Mucin and cadherin-like		5.1
PCDH9	Protocadherin 9		4.9
L1CAM	L1 cell adhesion molecule		4.2

*Overexpressed (E) or underexpressed (I) in adenomas (versus normal mucosa samples).

(30-32); decreased expression of the netrin-1 receptor, UNC5C (33); and expression changes involving three Fas apoptosis inhibitory molecules (FAIM), including FAIM1, which was increased 2.3-fold and is thus not listed in Table 2]; and (e) marked down-regulation of several genes that would result in reduced tumor suppression activity [e.g., those encoding the antiangiogenic factor ANGPTL1 (34), the cyclin-dependent kinase inhibitor CDKN2B/p15, and the prostaglandin catabolism enzyme HPGD (35)].

It is also important to recall the size-related differences noted in the adenoma gene expression profiles (Fig. 2; Supplementary Table S3). When validated in a larger series of tumors, these differences should provide important clues to the molecular basis of the well-known link between the dimensions and malignant potential of colorectal adenomas (1).

Our study also furnishes a complete picture of expression changes involving gene components of the Wnt pathway across the transition from normal to adenomatous epithelium (Supplementary Table S2) as well as evidence for the existence of a novel Wnt target: KIAA1199. This gene, which encodes a protein of unknown function, was strikingly overexpressed in all the adenomas included in this study and in 25 adenocarcinomas of the colon described in a previous report (8). Even more intriguingly, its expression was significantly correlated with that of several genes that are well-established targets of Wnt signaling. Our hypothesis that KIAA1199 is up-regulated by the TCF(s)/h-catenin transcription complex was considerably strengthened by the marked decreases in KIAA1199 expression observed in cultured colorectal cancer cells when the Wnt pathway was inhibited by overexpression of dominant-negative TCF4 proteins or by h-catenin knockdown. It is not yet clear whether this is a direct effect, but this possibility is supported by the results of a recent genome-wide TCF4 ChIP-on-chip analysis, which indicates that the KIAA1199 locus is surrounded by four TCF4-bound regions.[10] These findings are consistent with the probable role of this gene as a direct target of TCF4/h-catenin signaling in the intestine and in colorectal tumors.

Other features of KIAA1199 expression are also compatible with its putative role as a Wnt target gene. KIAA1199 mRNA and protein are both confined to the proliferative compartment of normal intestinal crypts, where Wnt signaling is normally active, and they are highly overexpressed in colorectal adenomas and carcinomas, where this pathway is almost always aberrantly activated.

In normal and tumor tissues, KIAA1199 is expressed in the cytoplasm of epithelial cells. In glands with low-degree dysplasia, higher concentrations are observed in the mucin vacuoles of goblet cells, but cytoplasmic expression of the protein in tumor cells remains elevated even after goblet cell differentiation has been lost (Fig. 4). These features, together with the localization of KIAA1199 in the luminal portion of the cytoplasm, are suggestive of a secreted and/or membrane protein. This conclusion is consistent with our in silico analysis of KIAA1199 (see Supplementary Data and Supplementary Fig. S5), which strongly predicts the presence of a signal peptide at its NH_2-terminal end. In addition, the central region of KIAA1199 contains a TMEM2 homology domain, which is present in several eukaryotic proteins, including TMEM2, polyductin (PKHD1), and fibrocystin L (PKHD1L1; Fig. 5), all large receptor proteins characterized by an NH_2-terminal signal peptide or a single transmembrane helix and a short cytoplasmic tail (36).

A study based on yeast two-hybrid screens suggested that KIAA1199 may interact with plexin A2 (KIAA0463; ref. 37). The transmembrane plexins interact with transmembrane semaphorins on nearby cells, providing "stop" and "go" signals that are crucial for cell motility and invasive growth (38, 39). KIAA1199/plexin A2 interaction could thus play important roles in colorectal tumorigenesis not only in the invasive stages but also earlier during the formation of abnormal glands in benign adenomas.

A recent report linked high levels of KIAA1199 mRNA with cell mortality in human fibroblasts and in a renal cell carcinoma cell line (40). In that study, however, there was no significant increase in KIAA1199 expression during replicative aging of mortal cells, and this finding contrasts with the documented behavior of other genes involved in cell aging (41). Furthermore,

[10] Hatzis et al., unpublished data.

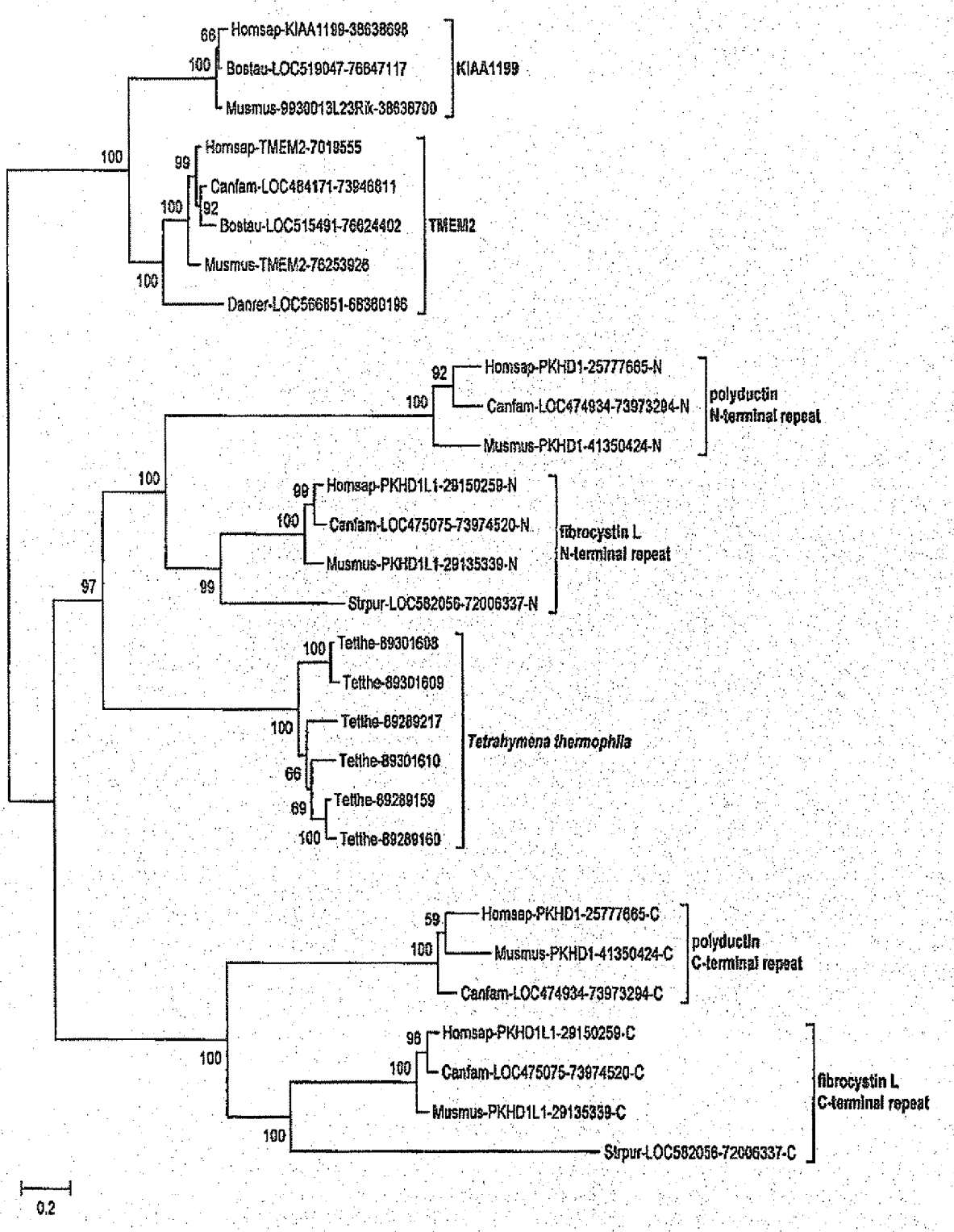


FIGURE 5. Phylogenetic tree of the proteins containing the TMEM2 homology domain found in the central region of KIAA1199. The tree was generated with MEGA3 (52) from the multiple sequence alignment shown in Supplementary Fig. S5. It was calculated with the minimum evolution algorithm and the JTT matrix. Positions with gaps were removed for calculation of pairwise distances. Node robustness was assessed using the bootstrap method with 100 resamplings. (Bootstrap values are shown at the nodes.) Two branches emerged, one comprising KIAA1199 and TMEM2 and the other with polyductin, fibrocystin L, and several other THD-containing proteins found in the ciliate Tetrahymena thermophila, which were apparently generated in a series of Tetrahymena-specific gene duplications. The NH_2-terminal repeats of polyductin and fibrocystin L clustered together, as did the COOH-terminal repeats, suggesting that the intragenic duplication of the TH domain in the ancestor of polyductin and fibrocystin L occurred before the divergence of chordates and echinoderms (more details in Supplementary Data).

the authors reported wide variation in KIAA1199 mRNA expression in breast cancer cell lines, and this finding raises the possibility that expression of this gene in vivo and in cell lines may differ.

We believe that our microarray data will serve as a springboard and reference point for other studies on the molecular basis of colorectal transformation along the adenoma-carcinoma pathway (and subsequently for the study of alternative pathways). Some of the transcriptional changes reported in this study might one day be used as molecular indices of the susceptibility of adenomas to malignant transformation, information that would be helpful in planning appropriate follow-up of the lesions. As for KIAA1199, its invariably high expression in the colorectal tumors we studied raises interesting possibilities for the development of a new molecular marker for the detection of these neoplasms. For example, because KIAA1199 expression in the normal mucosa is limited to cells in the lower portion of the crypts, which are not yet programmed to be shed into the intestinal lumen, the presence of KIAA1199 peptides in fecal water might prove to be a specific marker of adenomatous lesions. In addition, although due consideration must be given to its probable physiologic role(s) in intestinal crypts and possibly in several other human tissues (40, 42, 43), KIAA1199 may be a potential target of antibody-based therapies.

Materials and Methods

Tumor Samples

Pedunculated colorectal polyps and normal mucosa were obtained during colonoscopies carried out in the Gastroenterology Unit of the Belcolle City Hospital (Viterbo, Italy). The tissues were collected prospectively with informed patient consent and the approval of the local Human Research Ethics Committee. Patients with documented familial polyposis, with >15 adenomatous polyps (total: synchronous + previously excised; ref. 44), or currently treated with nonsteroidal anti-inflammatory drugs (including aspirin) were excluded from the study.

For each polyp, three biopsies of normal mucosa were collected from the same colon segment (≥2 cm from the site of the polyp). Immediately after removal, a small sample of epithelial tissue (5-15 mg) was cut from the tip of each polyp, leaving the underlying muscularis mucosae intact. We excluded polyps <1 cm to ensure that the sampling procedure would not interfere with the histologic diagnosis. All polyp samples were collected by a single operator (M.d.P.) using the same procedure to minimize artifacts due to sampling differences. The approach used allowed us to obtain specimens with a high percentage of epithelial cells without resorting to microdissection, which can diminish the quantity and quality of the extracted RNA.

The polyp sample and the three normal mucosal biopsies were immersed in RNAlater (Ambion) for subsequent microarray analysis, and the remainder of the polyp was submitted for pathologic analysis. The cut surface at the tip was labeled with India ink so that the sampled area could be easily identified during routine histologic examination. The tissue was then fixed in buffered formalin and embedded in paraffin. DNA extracted from sections of this specimen was also used to rule out microsatellite instability (reflecting defective DNA mismatch repair) at the BAT26 locus, as previously described (45). All of the polyps included in the study met the following criteria: type 0-Ip (6), maximum diameter of 1 to 4 cm, absence of surface ulceration, histologic diagnosis of adenoma, and absence of microsatellite instability at BAT26.

In some analyses, we also included transcriptomic data from a previously described set of 25 colon cancers (mismatch repair proficient and deficient; ref. 8), which we reanalyzed for this study with the same microarray used to characterize the adenomas and normal mucosa.

Microarray Analysis, Real-time Reverse Transcription-PCR, and Northern Blotting

Total RNA was extracted (RNeasy Mini kit, Qiagen) from homogenized tissue samples (5-15 mg), and its integrity was verified by capillary gel electrophoresis (Bio Analyzer, Agilent Technologies). Complementary RNA (15 Ag/sample), synthesized and labeled as previously described (8, 46), was hybridized with the Affymetrix U133 Plus 2.0 array, which contains in situ synthesized oligonucleotides representing the entire human genome (54,675 probes).

Raw gene expression data generated by GeneChip Operating Software (Affymetrix) were imported into the GeneSpring software program (Agilent Technologies) and normalized per chip (i.e., to the median of all values on a given array) and per gene (i.e., to the median expression level of the given gene across all samples). Analysis was done using the log expression values with GeneSpring's cross-gene error model turned on. Probes were excluded from analysis unless they were listed as "present or marginal calls" and/or had expression values ≥100 in ≥50% (≥16 of 32) of the samples in at least one of the tissue groups (adenomas and normal mucosa).

Expression data were subjected to four different unsupervised analyses: (a) hierarchical clustering using the Pearson correlation coefficient as a similarity measure and the average linkage algorithm for branch merging; (b) PCA, which reduces the dimensionality (number of variables) of a data set while retaining most of its variance (8); (c) correlation analysis, which involved computation of Pearson correlation coefficients for all possible sample pairs and visualization of correlation values as tile plots; and (d) CA, another dimension-reducing method (47), which was used to identify samples associated with particular gene expression levels. In typical CA, a matrix of n gene expression levels from p samples is treated as a two-way contingency table (genes by samples or vice versa) with n and p specifications for the "factors" gene and sample, respectively. Each intensity value thus reflects the abundance of a given transcript in a given sample. Like PCA, CA identifies independent "factorial components" that account for variance within a multidimensional gene data set, but in this case, the components are identified and ranked according to the correlation between gene and sample scores. A supervised or constrained extension of CA (9), CCA, was then used to identify possible correlations between gene expression patterns and clinical or pathologic variables. CA and CCA, as well as the corresponding plots, were computed using R software and the ade4 and made4 packages furnished by Bioconductor.[11]

[11] http://www.bioconductor.org

The Mann-Whitney test was used to select genes differentially expressed in normal mucosa and adenomas; Benjamini-Hochberg multiple testing correction was applied with a false discovery rate of 0.01. The genes in this set that were differentially expressed with fold differences of ≥2.0 were then analyzed with EmineJ software (48) to identify any biological processes from the Gene Ontology database (49) that were overrepresented.

Pearson correlation was used to identify correlation between KIAA1199 expression and the expression of other genes in the entire set of tissue samples. Fisher's exact test was used to identify possible overrepresentation of known Wnt targets among genes whose expression was closely correlated with that of KIAA1199 (correlation values ≥0.8).

Reverse transcription-PCR and Northern blotting were done as previously described (46, 50) to verify the expression level of KIAA1199 in tissue samples and in LS174T colon cancer cells in which inducible inhibition of the Wnt pathway had been achieved with previously described methods (14-16).

In situ Hybridization

Digoxigenin-labeled KIAA1199 antisense riboprobes were synthesized from a PCR product amplified from human colon cDNA with KIAA1199-specific primers (sense: 5]-cacatcggggaggagataga-3]; antisense, containing a T7 RNA polymerase-binding site: 5]-taatacgactcactatagggttccagactgaca-3]). This product was transcribed in vitro using the DIG RNA labeling kit and T7 RNA polymerase (Roche Diagnostics). In situ hybridizations were done on paraffin-embedded sections of human colon fixed with 4% buffered formalin as described elsewhere (51).

Immunohistochemistry

Our in silico analysis of KIAA1199 (see Supplementary Data) indicated that residues 202 to 217 (IHSDRFDTYRSKKESE) form a loop between a conserved h-strand and the following helix of the NH_2-terminal GG domain. This charged, surface-exposed peptide was used to raise a rabbit polyclonal antibody, which was purified by affinity chromatography on Thiopropyl Sepharose 6B (Amersham) derivatized with the antigenic peptide. A 1:1,000 dilution of this antibody was used, as previously described (45), to evaluate KIAA1199 expression in formalin-fixed, paraffin-embedded sections of adenoma and normal mucosal tissues.

Acknowledgments

We thank the patients who participated in this study; R. Haider, E. Pani, T. Lehmann, A. Patrignani, and U. Wagner for technical assistance; A. Mueller, E. Veljkovic, and F. Bannwart for critical reading of the manuscript; and M. Kent for editorial assistance.

References

1. Winawer SJ, Zauber AG, Fletcher RH, et al. Guidelines for colonoscopy surveillance after polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer and the American Cancer Society. Gastroenterology 2006;130:1872–85.

2. Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM, MacCarty RL. Natural history of untreated colonic polyps. Gastroenterology 1987;93:1009–13.

3. Pinto D, Clevers H. Wnt, stem cells and cancer in the intestine. Biol Cell 2005; 97:185–96.

4. Korinek V, Barker N, Morin PJ, et al. Constitutive transcriptional activation by a h-catenin-Tcf complex in APC-/- colon carcinoma. Science 1997;275: 1784–7.

5. Morin PJ, Sparks AB, Korinek V, et al. Activation of h-catenin-Tcf signaling in colon cancer by mutations in h-catenin or APC. Science 1997;275:1787–90.

6. Update on the Paris classification of superficial neoplastic lesions in the digestive tract. Endoscopy 2005;37:570–8.

7. Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology 2007;50:113–30.

8. di Pietro M, Sabates Bellver J, Menigatti M, et al. Defective DNA mismatch repair determines a characteristic transcriptional profile in proximal colon cancers. Gastroenterology 2005;129:1047–59.

9. Ter Braak C. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology 1986;67:1167–79.

10. Miki R, Kadota K, Bono H, et al. Delineating developmental and metabolic pathways in vivo by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays. Proc Natl Acad Sci U S A 2001;98:2199–204.

11. Wu LF, Hughes TR, Davierwala AP, Robinson MD, Stoughton R, Altschuler SJ. Large-scale prediction of Saccharomyces cerevisiae gene function using overlapping transcriptional clusters. Nat Genet 2002;31:255–65.

12. Zhang W, Morris QD, Chang R, et al. The functional landscape of mouse gene expression. J Biol 2004;3:21.

13. Van der Flier LG, Sabates-Bellver J, Oving I, et al. The intestinal Wnt/TCF signature. Gastroenterology 2007;132:628–32.

14. van de Wetering M, Barker N, Harkes IC, et al. Mutant E-cadherin breast cancer cells do not display constitutive Wnt signaling. Cancer Res 2001;61: 278–84.

15. van de Wetering M, Sancho E, Verweij C, et al. The h-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell 2002;111:241–50.

16. van de Wetering M, Oving I, Muncan V, et al. Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector. EMBO Rep 2003;4:609–15.

17. Notterman DA, Alon U, Sierk AJ, Levine AJ. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res 2001;61:3124–30.

18. Lin YM, Furukawa Y, Tsunoda T, Yue CT, Yang KC, Nakamura Y. Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. Oncogene 2002;21: 4120–8.

19. Nosho K, Yamamoto H, Adachi Y, Endo T, Hinoda Y, Imai K. Gene expression profiling of colorectal adenomas and early invasive carcinomas by cDNA array analysis. Br J Cancer 2005;92:1193–200.

20. Lechner S, Muller-Ladner U, Renke B, Scholmerich J, Ruschoff J, Kullmann F. Gene expression pattern of laser microdissected colonic crypts of adenomas with low grade dysplasia. Gut 2003;52:1148–53.

21. Nata K, Liu Y, Xu L, et al. Molecular cloning, expression and chromosomal localization of a novel human REG family gene, REG III. Gene 2004;340:161–70.

22. Bishnupuri KS, Luo Q, Murmu N, Houchen CW, Anant S, Dieckgraefe BK. Reg IV activates the epidermal growth factor receptor/Akt/AP-1 signaling pathway in colon adenocarcinomas. Gastroenterology 2006;130:137–49.

23. Hanai J, Mammoto T, Seth P, et al. Lipocalin 2 diminishes invasiveness and metastasis of Ras-transformed cells. J Biol Chem 2005;280:13641–7.

24. Comijn J, Berx G, Vermassen P, et al. The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. Mol Cell 2001;7:1267–78.

25. Jankowski JA, Bedford FK, Boulton RA, et al. Alterations in classical cadherins associated with progression in ulcerative and Crohn's colitis. Lab Invest 1998;78:1155–67.

26. Hardy RG, Tselepis C, Hoyland J, et al. Aberrant P-cadherin expression is an early event in hyperplastic and dysplastic transformation in the colon. Gut 2002; 50:513–9.

27. Lin SY, Elledge SJ. Multiple tumor suppressor pathways negatively regulate telomerase. Cell 2003;113:881–9.

28. Mariadason JM, Nicholas C, L'Italien KE, et al. Gene expression profiling of intestinal epithelial cell maturation along the crypt-villus axis. Gastroenterology 2005;128:1081–8.

29. Ding H, Schertzer M, Wu X, et al. Regulation of murine telomere length by Rtel: an essential gene encoding a helicase-like protein. Cell 2004;117:873–86.

30. Pitti RM, Marsters SA, Lawrence DA, et al. Genomic amplification of a

decoy receptor for Fas ligand in lung and colon cancer. Nature 1998;396:699–703.

31. Bai C, Connolly B, Metzker ML, et al. Overexpression of M68/DcR3 in human gastrointestinal tract tumors independent of gene amplification and its location in a four-gene cluster. Proc Natl Acad Sci U S A 2000;97:1230–5.

32. Postma C, Hermsen MA, Coffa J, et al. Chromosomal instability in flat adenomas and carcinomas of the colon. J Pathol 2005;205:514–21.

33. Thiebault K, Mazelin L, Pays L, et al. The netrin-1 receptors UNC5H are putative tumor suppressors controlling cell death commitment. Proc Natl Acad Sci U S A 2003;100:4173–8.

34. Dhanabal M, Jeffers M, LaRochelle WJ, Lichenstein HS. Angioarrestin: a unique angiopoietin-related protein with anti-angiogenic properties. Biochem Biophys Res Commun 2005;333:308–15.

35. Backlund MG, Mann JR, Holla VR, et al. 15-Hydroxyprostaglandin dehydrogenase is down-regulated in colorectal cancer. J Biol Chem 2005;280:3217–23.

36. Hogan MC, Griffin MD, Rossetti S, Torres VE, Ward CJ, Harris PC. PKHDL1, a homolog of the autosomal recessive polycystic kidney disease gene, encodes a receptor with inducible T lymphocyte expression. Hum Mol Genet 2003;12:685–98.

37. Nakayama M, Kikuno R, Ohara O. Protein-protein interactions between large proteins: two-hybrid screening using a functionally classified library composed of long cDNAs. Genome Res 2002;12:1773–84.

38. Tamagnone L, Artigiani S, Chen H, et al. Plexins are a large family of receptors for transmembrane, secreted, and GPI-anchored semaphorins in vertebrates. Cell 1999;99:71–80.

39. Comoglio PM, Tamagnone L, Giordano S. Invasive growth: a two-way street for semaphorin signalling. Nat Cell Biol 2004;6:1155–7.

40. Michishita E, Garces G, Barrett JC, Horikawa I. Upregulation of the KIAA1199 gene is associated with cellular mortality. Cancer Lett 2006;239:71–7.

41. Horikawa I, Parker ES, Solomon GG, Barrett JC. Upregulation of the gene encoding a cytoplasmic dynein intermediate chain in senescent human cells. J Cell Biochem 2001;82:415–21.

42. Abe S, Katagiri T, Saito-Hisaminato A, et al. Identification of CRYM as a candidate responsible for nonsyndromic deafness, through cDNA microarray analysis of human cochlear and vestibular tissues. Am J Hum Genet 2003;72:73–82.

43. Abe S, Usami S, Nakamura Y. Mutations in the gene encoding KIAA1199 protein, an inner-ear protein expressed in Deiters' cells and the fibrocytes, as the cause of nonsyndromic hearing loss. J Hum Genet 2003;48:564–70.

44. Marra G, Jiricny J. Multiple colorectal adenomas—is their number up? N Engl J Med 2003;348:845–7.

45. Truninger K, Menigatti M, Luz J, et al. Immunohistochemical analysis reveals high frequency of PMS2 defects in colorectal cancer. Gastroenterology 2005;128:1160–71.

46. di Pietro M, Marra G, Cejka P, et al. Mismatch repair-dependent transcriptome changes in human cells treated with the methylating agent N-methyl-N'-nitro-N-nitrosoguanidine. Cancer Res 2003;63:8158–66.

47. Benzécri J-P. Correspondence analysis handbook. New York: Marcel Dekker, Inc.; 1992.

48. Lee HK, Braynen W, Keshav K, Pavlidis P. ErmineJ: tool for functional analysis of gene expression data sets. BMC Bioinformatics 2005;6:269.

49. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–9.

50. Barker N, Hurlstone A, Musisi H, Miles A, Bienz M, Clevers H. The chromatin remodelling factor Brg-1 interacts with β-catenin to promote target gene activation. EMBO J 2001;20:4935–43.

51. Meake C, Auf der Maur F, Jovanovic K, Reinecke M, Hauri D, John H. Occurrence and localization of uroguanylin in the aging human prostate. Histochem Cell Biol 2003;119:69–76.

52. Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 2004;5:150–63.